

**UNIVERSIDADE DO ESTADO DO AMAZONAS  
ESCOLA SUPERIOR DE TECNOLOGIA  
ENGENHARIA ELÉTRICA**

**ANA CLARA BEGNINI**

**PREDIÇÃO DE CLASSES SOCIAIS COM MODELOS DE  
APRENDIZADO PROFUNDO A PARTIR DE IMAGENS DE SATÉLITE  
E DADOS DE RENDA DO CENSO**

Manaus  
2023

ANA CLARA BEGNINI

**PREDIÇÃO DE CLASSES SOCIAIS COM MODELOS DE  
APRENDIZADO PROFUNDO A PARTIR DE IMAGENS DE SATÉLITE  
E DADOS DE RENDA DO CENSO**

Pesquisa desenvolvida durante a disciplina de Trabalho de Conclusão de Curso II e apresentada à banca avaliadora do Curso de Engenharia Elétrica da Escola Superior de Tecnologia da Universidade do Estado do Amazonas, como pré-requisito para obtenção do título de Engenheiro Eletricista.

Orientador: Prof. Dr. Carlos Maurício Serodio Figueiredo

Manaus

2023

**Universidade do Estado do Amazonas – UEA**  
**Escola Superior de Tecnologia - EST**

Reitor:

**André Luiz Nunes Zodaib**

Vice-Reitor:

**Kátia do Nascimento Coureiro**

Diretora da Escola Superior de Tecnologia:

**Ingrid Sammyne Gadelha Figueiredo**

Coordenador do Curso de Engenharia Elétrica:

**Israel Gondres Torné**

Banca Avaliadora composta por: Data da defesa: <27/03/2023>.

**Prof. Carlos Maurício Serodio Figueiredo, Dr. (Orientador)**

**Prof. Antonio Luiz Alencar Pantoja, Dr.**

**Prof. Jozias Parente de Oliveira, Dr.**

## **CIP – Catalogação na Publicação**

Begnini, Ana Clara

Predição de classes sociais com modelos de aprendizado profundo a partir de imagens de satélite e dados de renda do Censo / Ana Clara Begnini; [orientado por] Carlos Maurício Serodio Figueiredo. – Manaus: 2023.

69 p.: il.

Trabalho de Conclusão de Curso (Graduação em Engenharia Elétrica). Universidade do Estado do Amazonas, 2023.

1. Imagens de satélite. 2. Dados de renda. 3. Deep Learning.  
I. Figueiredo, Carlos Maurício Serodio.

**ANA CLARA BEGNINI**

**PREDIÇÃO DE CLASSES SOCIAIS COM MODELOS DE APRENDIZADO  
PROFUNDO A PARTIR DE IMAGENS DE SATÉLITE E DADOS DE RENDA DO  
CENSO**

Pesquisa desenvolvida durante a disciplina de Trabalho de Conclusão de Curso II e apresentada à banca avaliadora do Curso de Engenharia Elétrica da Escola Superior de Tecnologia da Universidade do Estado do Amazonas, como pré-requisito para a obtenção do título de Engenheiro Eletricista.

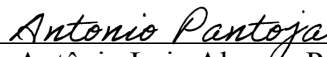
Nota obtida: 10,0 (dez vírgula zero)


Aprovada em 27/03/2023

Área de concentração: Inteligência Artificial

**BANCA EXAMINADORA**

  
Orientador: Carlos Maurício Serodio Figueiredo, Dr.

  
Avaliador: Antônio Luiz Alencar Pantoja, Dr.

  
Avaliador: Joziás Parente de Oliveira, Dr.

Manaus  
2023



## **Dedicatória**

À minha mãe, Rosângela Begnini, que sempre me ensinou sobre ter força, determinação e foco em tudo aquilo que faço. Por todo o apoio em minha formação, tanto profissional quanto pessoal, e por sempre estar presente nas minhas conquistas, afinal, são nossas conquistas, porque sem a senhora não seria possível.

## AGRADECIMENTO

Agradeço aos meus pais, Rosângela e Régis Begnini, por sempre estarem me incentivando e apoiando em meus estudos, além de estarem presente nos momentos felizes e difíceis da minha vida dando todo suporte que preciso para continuar.

Agradeço ao meu orientador, Carlos Maurício, por ter me orientado e acompanhado grande parte da minha trajetória acadêmica, sempre me incentivando a dar meu melhor nos projetos e nunca desistir.

Agradeço à minha maior referência e inspiração feminina na área que trabalho atualmente, Alice Adativa, por tudo que me ensinou, sempre me ajudando, auxiliando e incentivando nos projetos e trabalhos que fizemos juntas.

Agradeço à todos os amigos e colegas que reencontrei e conheci nesse período na EST. Ao meu companheiro e parceiro de jornada, Vinícius dos Santos, por sempre me lembrar, quando preciso, que sou capaz de realizar todos os meus sonhos. Às minhas melhores amigas, Isabel Priscila e Talissa Moura, por estarem sempre ao meu lado nos momentos difíceis e nos momentos de conquistas. Aos meus amigos, Holyvers Vinícius, Marcus Cunha e Rubem Silas, que caminharam junto comigo no desafio que foi cursar Engenharia Elétrica. E aos meus amigos, Josué Alves, Gabriel Rodrigues, Paulo Henrique e Paulo Victor, por todas as saídas que fizemos para comemorar as aprovações e as conquistas, além de aliviar as tensões durante um período complicado.

Agradeço à essas e todas as pessoas que me ensinaram, ajudaram e apoiaram de alguma forma nessa fase desafiadora que foi cursar Engenharia.

## RESUMO

As pesquisas socioeconômicas que coletam informações sobre renda ou situação econômica das famílias no território brasileiro são pesquisas custosas e que demandam tempo para se realizar, por exemplo, o Censo Demográfico brasileiro são realizadas a cada 10 anos, no mínimo, o que pode tornar as formulações de políticas públicas não eficientes visto que, nos últimos anos, é notório um crescimento populacional significativo nas cidades. Contudo, o presente trabalho tem o objetivo de estudar e analisar a possibilidade de classificar as áreas das cidades de Manaus (AM) e São Paulo (SP) entre as classes sociais por meio de modelos de aprendizado profundo com imagens de satélite das cidades e dados de renda do último Censo Demográfico disponível. Para isso, foram coletadas as imagens de satélites de cada cidade e, a partir de dados de renda do Censo 2010, classificá-las entre as classes sociais A, B, C, D e E, com isso, foi treinado modelos de Visão Computacional para problemas de classificação com a arquitetura *EfficientNetV2* utilizando a técnica do *transfer learning* para o treinamento. Por fim, obteve-se 16 modelos de aprendizado profundo sendo 8 da cidade de Manaus e 8 da cidade de São Paulo, porém, os 2 melhores modelos de Manaus obtiveram o F1-Score de 0,66 para o Modelo 1 e 0,48 para o Modelo 2, enquanto que os 2 melhores modelos de São Paulo obtiveram resultados de F1-Score 0,53 para o Modelo 1 e 0,58 para o Modelo 2.

**Palavras chave:** Imagens de satélite, dados de renda, *Deep Learning*, classes sociais.

## ABSTRACT

Socioeconomic surveys that collect information on the income or economic situation of families in the Brazilian territory are costly surveys that demand time to be carried out, for example, the Brazilian Demographic Census is carried out every 10 years, at least, which can make the formulations of inefficient public policies since, in recent years, significant population growth in cities has been notorious. However, the present work has the objective of studying and analyzing the possibility of classifying the areas of the cities of Manaus (AM) and São Paulo (SP) according to social classes through deep learning models with satellite images of the cities and data from the last available Population Census. For this, satellite images of each city were collected and, based on income data from the 2010 Census, classified into social classes A, B, C, D, and E, with this, Computational Vision models were trained for classification problems with the EfficientNetV2 architecture using the transfer learning technique for training. Finally, it obtained 16 deep learning models, 8 from the city of Manaus and 8 from the city of São Paulo, however, the 2 best models from Manaus obtained the F1-Score of 0.66 for Model 1 and 0.48 for Model 2, while the 2 best models from São Paulo obtained F1-Score results of 0.53 for Model 1 and 0.58 for Model 2.

**Keywords:** Satellite images, income data, Deep Learning, social classes.

## Lista de Figuras

1	Representação do Deep Learning . . . . .	21
2	Diagrama com a operação de convolução 2-D . . . . .	22
3	Exemplo de aprendizado das CNNs . . . . .	23
4	Operação de Convolução em Imagens . . . . .	24
5	Objetivo Geral de modelos de classificação . . . . .	25
6	Tempo de Treino vs Acurácia . . . . .	26
7	OpenCV Logo . . . . .	29
8	TensorFlow Logo . . . . .	30
9	Keras Logo . . . . .	30
10	Diagrama das Etapas do Desenvolvimento do Trabalho . . . . .	31
11	Setores censitários de cada estado . . . . .	32
12	Setores censitários de cada cidade . . . . .	33
13	Limites máximo e mínimo de cada cidade no formato de um retângulo . . . . .	33
14	Total de <i>Grids</i> envolvendo todo o retângulo de cada cidade . . . . .	34
15	<i>Grids</i> dentro dos limites dos setores censitários em cada cidade . . . . .	35
16	Exemplo das imagens coletadas de ambas as cidades . . . . .	36
17	Exemplo da Interseção de 1 <i>grid</i> com 2 setores censitários . . . . .	40
18	Diagrama para a rotulação das imagens para cada cidade . . . . .	41
19	Diagrama de Treinamento . . . . .	42
20	<i>Pipeline</i> das transformações no aumento de dados de imagens . . . . .	43
21	Matrizes de confusão dos modelos da cidade de Manaus - Treinamento com o aumento de dados e pesos das categorias . . . . .	47
22	Matrizes de confusão dos modelos da cidade de Manaus - Treinamento com o conjunto de dados balanceado . . . . .	48
23	Matrizes de confusão dos modelos da cidade de São Paulo - Treinamento com o aumento de dados e pesos das categorias . . . . .	49
24	Matrizes de confusão dos modelos da cidade de São Paulo - Treinamento com o conjunto de dados balanceado . . . . .	50

25	Histogramas das classes sociais reais x das classes sociais previstas pelo modelo na base de teste - Manaus . . . . .	56
26	Histogramas com as distribuições de probabilidade para cada classe social do Modelo 1 da cidade de Manaus . . . . .	57
27	Histogramas com as distribuições de probabilidade para cada classe social do Modelo 2 da cidade de Manaus . . . . .	59
28	Histogramas das classes sociais reais x das classes sociais previstas pelo modelo na base de teste - São Paulo . . . . .	61
29	Histogramas com as distribuições de probabilidade para cada classe social do Modelo 1 da cidade de São Paulo . . . . .	62
30	Histogramas com as distribuições de probabilidade para cada classe social do Modelo 2 da cidade de São Paulo . . . . .	64

## Lista de Tabelas

1	Matriz de Confusão . . . . .	27
2	Informações selecionadas na tabela de Pessoa Renda . . . . .	37
3	Informações selecionadas na tabela de Responsável Renda . . . . .	38
4	Distribuição das classes sociais nos conjuntos de dados . . . . .	43
5	Distribuição das classes sociais nos conjuntos de dados de treino . . . . .	44
6	Distribuição das classes sociais nos conjuntos de dados balanceados . . . . .	45
7	Métricas de avaliação dos modelos de Manaus: Treinamento com o aumento de dados e pesos das categorias . . . . .	51
8	Métricas de avaliação dos modelos de Manaus: Treinamento com as bases balanceadas . . . . .	52
9	Métricas de avaliação dos modelos de São Paulo: Treinamento com o aumento de dados e pesos das categorias . . . . .	53
10	Métricas de avaliação dos modelos de São Paulo: Treinamento com as bases balanceadas . . . . .	54
11	Erro Absoluto do Modelo 1 da cidade de Manaus . . . . .	58
12	Erro Absoluto do Modelo 2 da cidade de Manaus . . . . .	60
13	Erro Absoluto do Modelo 1 da cidade de São Paulo . . . . .	63
14	Erro Absoluto do Modelo 2 da cidade de São Paulo . . . . .	65

## Lista de Abreviatura

AM	Amazonas
API	<i>Application Programming Interface</i>
CNN	<i>Convolutional Neural Networks</i>
CPU	<i>Central Processing Units</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FPGA	<i>Field Programmable Gate Array</i>
FPN	<i>Feature Pyramid Network</i>
GPU	<i>Graphic Processing Units</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
MAE	<i>Mean Absolute Error</i>
OpenCV	<i>Open Source Computer Vision Library</i>
P	Precisão
PIB	Produto Interno Bruto
R	Revocação
RELU	<i>Rectified Linear Unit</i>
ResNet	<i>Residual Network</i>
RoI	Região de Interesse
SP	São Paulo
SR	Sensoriamento Remoto
TN	<i>True Negative</i>
TP	<i>True Positive</i>
TPU	<i>Tensor Processing Unit</i>
VGG	<i>Visual Geometry Group</i>



## SUMÁRIO

<b>INTRODUÇÃO</b>	<b>15</b>
<b>1 PROBLEMA DE PESQUISA</b>	<b>16</b>
<b>2 OBJETIVOS</b>	<b>16</b>
2.1 OBJETIVO GERAL	16
2.2 OBJETIVOS ESPECÍFICOS	16
<b>3 JUSTIFICATIVA</b>	<b>16</b>
<b>4 TRABALHOS RELACIONADOS</b>	<b>17</b>
<b>5 REFERENCIAL TEÓRICO</b>	<b>18</b>
5.1 CENSO DEMOGRÁFICO	18
5.1.1 SETORES CENSITÁRIOS	19
5.2 SENSORIAMENTO REMOTO	19
5.3 APRENDIZADO DE MÁQUINA	19
5.3.1 APRENDIZADO SUPERVISIONADO	20
5.4 APRENDIZADO PROFUNDO	20
5.4.1 REDES NEURAIIS CONVOLUCIONAIS	21
5.4.2 MODELOS DE CLASSIFICAÇÃO DE IMAGENS	24
5.4.3 TRANSFER LEARNING	26
5.4.4 MÉTRICAS DE VALIDAÇÃO DOS MODELOS	27
5.5 PYTHON	29
5.6 OPENCV	29
5.7 TENSORFLOW E KERAS	29
<b>6 METODOLOGIA</b>	<b>31</b>
6.1 CRIAÇÃO DO CONJUNTO DE DADOS	31
6.1.1 COLETA DAS IMAGENS DE SATÉLITES	32
6.1.2 DADOS DE RENDA DO CENSO DE 2010	36
6.1.3 PROCESSO DE ANOTAÇÃO DAS IMAGENS	39
6.2 PREPARAÇÃO DO CONJUNTO DE DADOS E TREINAMENTO DOS MODELOS	41
6.2.1 PREPARAÇÃO DO CONJUNTO DE DADOS PARA TREINAMENTO COM O AUMENTO DE DADOS E PESOS DAS CLASSES	43

6.2.2	PREPARAÇÃO DO CONJUNTO DE DADOS PARA TREINAMENTO COM A BASE BALANCEADA . . . . .	45
<b>7</b>	<b>RESULTADOS E DISCUSSÕES . . . . .</b>	<b>46</b>
7.1	MATRIZ DE CONFUSÃO . . . . .	46
7.2	PRECISÃO, REVOCAÇÃO E F1-SCORE . . . . .	51
7.3	ANÁLISE DOS MODELOS EM CADA CLASSE SOCIAL . . . . .	55
7.3.1	MODELOS SELECIONADOS DE MANAUS (AM) . . . . .	55
7.3.2	MODELOS SELECIONADOS DE SÃO PAULO (SP) . . . . .	60
<b>8</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>66</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>68</b>

## INTRODUÇÃO

Analisando os dados sobre a estimativa da população brasileira coletada pelo Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>1</sup> nos últimos 15 anos, é possível verificar um aumento de mais de 15% da população, o que pode ser considerado um crescimento significativo nos centros urbanos. Com esse crescimento acelerado, é preciso ferramentas e tecnologias com a finalidade de auxiliar no planejamento urbano e socioeconômico, caso contrário, é observado uma queda na qualidade de vida do ser humano e uma maior desigualdade social na cidade.

Existem algumas pesquisas socioeconômicas que coletam informações sobre a renda ou situação econômica das famílias de uma região, uma delas é o Censo Demográfico que coleta várias informações sobre a qualidade de vida do ser humano na cidade, dentre elas, a renda familiar. Porém, essas pesquisas socioeconômicas são custosas e é um processo que demanda muito tempo para realizar (CASTRO; ÁLVAREZ, 2022; SHEEHAN et al., 2019). No Brasil, as pesquisas do Censo são realizadas no período de 10 anos sendo que a última foi feita em 2010 e a próxima ocorreu em 2022 devido ao atraso em 2020 por conta do COVID-19 (IBGE, 2022), porém até o presente momento os dados ainda não foram divulgados e disponibilizados.

Esse longo prazo entre as pesquisas dificultam na elaboração das formulações adequadas das políticas públicas que visam um melhor planejamento urbano, uma melhora na qualidade de vida do ser humano e a diminuição das desigualdades sociais da cidade (CASTRO; ÁLVAREZ, 2022). Logo, nos últimos anos estão surgindo várias pesquisas que aceleram esse processo de obter dados socioeconômicos em diversos países, como o trabalho de Castro e Álvarez (2022) que utiliza imagens de satélites para obter indicadores socioeconômicos do Brasil, o trabalho de Sheehan et al. (2019) que utiliza artigos da *Wikipedia* e as imagens de satélites para estimar os indicadores socioeconômicos de alguns países da África e o trabalho de Fatehkia et al. (2020) que utiliza dados de publicidade coletados do *Facebook* para mapear os indicadores socioeconômicos da Filipinas e da Índia.

Com isso, o presente trabalho é um estudo para acelerar o processo de análise da distribuição e da proporção das classes sociais da cidade de Manaus (Amazonas) e da cidade de São Paulo (São Paulo) por meio de imagens de satélite e dados do Censo. Para isso, foram implementados modelos de Aprendizado Profundo com a técnica de *transfer learning* e, a partir dos resultados, examinados e avaliados os modelos quanto à precisão e à razoabilidade com a finalidade de determinar a eficácia de cada um.

---

<sup>1</sup><https://www.ibge.gov.br/>

## **1 PROBLEMA DE PESQUISA**

Censos e outras pesquisas responsáveis pela coleta de dados socioeconômicos são caros e demorados. No Brasil, essas pesquisas são realizadas em um período de 10 anos, o que dificulta na elaboração das formulações adequadas de políticas públicas a fim de ter um melhor planejamento urbano, uma melhor qualidade de vida do homem e uma diminuição das desigualdades sociais da cidade.

## **2 OBJETIVOS**

### **2.1 OBJETIVO GERAL**

Classificar as classes sociais das áreas urbanas da cidade de Manaus - AM e São Paulo -SP utilizando Imagens de Satélites, dados de renda do Censo e implementá-los nos modelos de Aprendizado Profundo para automatizar e acelerar o processo de estimar as classes sociais e, assim, auxiliar no planejamento urbano e monitoramento socioeconômico das cidades.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Criar um conjunto de dados de imagens de satélite rotuladas com os dados de renda coletados do Censo do IBGE para o contexto de pesquisa apresentada;
- Pesquisar e avaliar comparativamente diferentes modelos de Aprendizado Profundo utilizados para classificação que seja eficaz na predição a partir de imagens de satélite;
- Disponibilizar um modelo de Aprendizado Profundo para a estimativa de classes sociais a partir de imagens que possa ser usado como ferramenta auxiliar de monitoramento e planejamento urbano.

## **3 JUSTIFICATIVA**

Nos países mais humildes e em desenvolvimento, há uma longa distância entre as pesquisas socioeconômicas, principalmente do Censo, o que dificulta as formulações adequadas de políticas públicas. No Brasil, o Censo é realizado em 10 anos e é uma pesquisa custosa de se realizar visto que as informações e os resultados da pesquisa são coletados em todo território brasileiro. Com isso, o trabalho proposto se justifica por ser um meio de automatizar ou ao menos ter uma estimativa das classes sociais da cidade em um período de tempo menor que as pesquisas atuais, além de ser um problema que se

enquadra no tema de Cidades Inteligentes por conta do auxílio no planejamento urbano e monitoramento socioeconômico da cidade.

Além disso, a implementação de modelos convolucionais para solucionar o problema em questão, de maneira geral, envolve a utilização de vários conceitos estudados nas disciplinas de Engenharia Elétrica, tais como: Álgebra Linear I e II; Cálculo I, II, III e IV; Probabilidade e Estatística; Linguagem de Programação I; Processamento Digital de Imagens; Tópicos Especiais para Computação II.

#### 4 TRABALHOS RELACIONADOS

Nos últimos anos estão surgindo varias pesquisas que aceleram o processo de obter dados socioeconômicos em diversos países.

O trabalho "*Predicting socioeconomic indicators using transfer learning on imagery data: an application in Brazil*" dos autores Castro e Álvarez (2022) utiliza técnicas de aprendizado profundo com imagens de satélites para prever os indicadores socioeconômicos no Brasil, mas especificamente, nos estados da Bahia e do Rio Grande do Sul. Eles utilizam imagens coletadas tanto durante o dia quanto a noite para prever a renda média, o produto interno bruto (PIB) e índice de água por meio do método do *transfer learning* com a arquitetura de rede neural convolucional chamada VGG16 para realizar a extração de *features* (características) de cada imagem e, com as *features*, treinam outro modelo de rede neural convolucional para prever as características das luzes noturnas das cidades, que é outro ponto deste trabalho em questão. O trabalho de Castro e Álvarez (2022) se relaciona com o presente trabalho justamente porque utiliza os dados de renda e PIB do Censo de 2010 e também foi a base para determinar área das imagens de satélite coletadas.

Outro trabalho interessante é o "*Estimation of Poverty Based on Remote Sensing Image and Convolutional Neural Network*" dos autores Wu e Tan (2019) que também utiliza redes neurais convolucionais com o *transfer learning* para extrair as *features* das imagens de satélite com a finalidade de estimar o nível de pobreza na cidade da China chamada Guizhou. Eles utilizaram também imagens com luzes noturnas para classificar as imagens com relação ao índice de vegetação, índice de água e índice das construções urbanas. Para o classificador, utilizaram a arquitetura ResNet50 com as camadas residuais e uma rede de pirâmide de características (FPN). Como resultados, os autores afirmaram que essas *features* extraídas pelo aprendizado profundo podem explicar até 76% da variação nos dados econômicos locais.

Além desses, há também o trabalho "*Predicting City Poverty Using Satellite Imagery*" dos autores Piaggese et al. (2019) que também utiliza o processo do *transfer learning* com imagens de satélite para prever os indicadores socioeconômicos, neste caso sendo a

renda familiar, nas cidades Los Angeles, Chicago, Houston, Filadélfia e Boston dos Estados Unidos e na cidade Santiago do Chile. Realizaram o treinamento considerando 3 modelos de redes neurais convolucionais: um *transfer learning* com ResNet50, outro com o VGG e, por último, um *transfer learning com VGG* adicionando um *fine-tuning*, que é um ajuste fino do modelo, com os rótulos de intensidade de luzes noturnas. Compararam os resultados dos 3 modelos sendo o do ResNet50 obtendo o melhor desempenho na maioria dos casos.

## 5 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os conceitos e fundamentos teóricos que estão relacionados ao desenvolvimento do projeto proposto. Inicialmente na Seção 5.1 é abordado sobre a pesquisa do Censo que é realizado no Brasil, posteriormente, na Seção 5.2, é retratado sobre o conceito do sensoriamento remoto. Na seção 5.3 é abordado sobre aprendizado de máquina, enquanto que na Seção 5.4 é abordado sobre o aprendizado profundo, onde que na subseção 5.4.1 é direcionado para os fundamentos e conceitos das Redes Neurais Convolucionais, enquanto que na subseção 5.4.2 é focado nas arquiteturas que serão implementadas no projeto e nas subseções 5.4.3 e 5.4.4 é apresentada a técnica para o treinamento dos modelos que serão utilizados e as principais métricas de validação que serão utilizadas para avaliar os modelos desenvolvidos, respectivamente. Já na Seção 5.5 é abordado sobre a linguagem de programação que será utilizada para o desenvolvimento do projeto e, por fim, nas Seções 5.6 e 5.7 são apresentadas as bibliotecas que serão utilizadas para o processamento de imagens e para o desenvolvimento dos modelos das redes neurais, respectivamente.

### 5.1 CENSO DEMOGRÁFICO

De acordo com o IBGE (2022), o Censo Demográfico é a principal fonte de referência para o conhecimento da qualidade de vida da população em todos os municípios do País e em seus recortes internos, como distritos, bairros e localidades. Seu objetivo é coletar informações como número de habitantes do território nacional, identificação étnico-racial, fecundidade, deficiência, migração interna ou internacional, educação, deslocamento para estudo e trabalho, mortalidade, autismo, renda familiar, entre outras informações a fim de analisar as condições de vida da população. Com isso, é possível definir políticas públicas e tomadas de decisões de investimentos da iniciativa privada ou de qualquer nível de governo.

Em 1808, foi efetuado o primeiro recenseamento da população brasileira com a finalidade de atender os interesses militares. Porém, por efeito do registro histórico, o

recenseamento de 1872 (denominado Censo Geral do Império) é considerado o primeiro efetuado no País, cujo foi conduzido pela Directoria-Geral de Estatística (IBGE, 2022).

A periodicidade da pesquisa é a cada 10 anos. Entretanto, nos anos de 1910 e 1930 o levantamento foi suspenso, no ano de 1990 a operação foi adiada para 1991 e, no ano de 2020, a coleta foi postergada por conta da pandemia de COVID-19, logo a próxima coleta do Censo Demográfico será realizada entre os meses de agosto a outubro de 2022, no qual serão investigados cerca de 78 milhões de domicílios particulares permanentes do País (IBGE, 2022).

### 5.1.1 SETORES CENSITÁRIOS

O Setor Censitário é uma pequena região de terra que o IBGE utiliza para planejar, coletar e disseminar os resultados dos Censos e Pesquisas Estatísticas realizadas em cada estado nacional. A partir de 2019, foi separado o termo Malha Setorial Censitária da Malha Setorial Intermediária com o objetivo de separar os resultados de pesquisas que possuem um levantamento minucioso de todo o território brasileiro, como o Censo Demográfico e Censo Agropecuário, visto que o segundo termo é para as pesquisas anuais atualizadas com as mudanças territoriais, porém, a mesma não obtém informações estatísticas de cada setor como a Malha Setorial Censitária (IBGE, 2023).

## 5.2 SENSORIAMENTO REMOTO

De acordo com Elachi e Zyl (2006), o Sensoriamento Remoto (SR) é definido como uma aquisição de informações sobre um objeto sem estar em contato físico com o mesmo. Um dos principais objetivos do SR é a aquisição de dados sobre a superfície terrestre com a finalidade de mapear e avaliar os recursos terrestres, além do monitoramento ambiental (IBGE, 2001).

As informações são obtidas através de radiação eletromagnética, que pode ser gerada por fontes naturais, como o Sol, ou por fontes artificiais, como o radar, que são refletidas ou emitidas pelos objetos e capturados por sensores óticos orbitais localizados em satélites. Essas informações são apresentadas em forma de imagens. Ou seja, o processo envolve a existência de uma fonte de emissão de radiação eletromagnética e um sensor para a captação do sinal após ser incidir sobre o alvo (IBGE, 2001).

## 5.3 APRENDIZADO DE MÁQUINA

De acordo com Goodfellow, Bengio e Courville (2015), o algoritmo de Aprendizado de Máquina (*"Machine Learning"*) é um algoritmo capaz de aprender a partir de dados.

Uma das tarefas mais comuns em *machine learning* são os problemas de classificação, onde que o programa é solicitado a especificar qual das  $k$  categorias alguma entrada pertence, para isso, o algoritmo a função:

$$f : \mathbb{R}^n \rightarrow \{1, \dots, k\} \quad (1)$$

Logo,  $y = f(x)$  é quando o modelo associou as entradas descritas por vetor  $x$  para uma categoria identificada por  $y$ . Existem algumas formas de realizar o treinamento dos algoritmos (processo em que o mesmo aprende através dos dados), sendo elas: aprendizado supervisionado, aprendizado não-supervisionado, aprendizado semi-supervisionado, aprendizado auto-supervisionado e aprendizado por reforço. No caso do desenvolvimento deste trabalho (CHOLLET, 2018), foi utilizado o aprendizado supervisionado.

### 5.3.1 APRENDIZADO SUPERVISIONADO

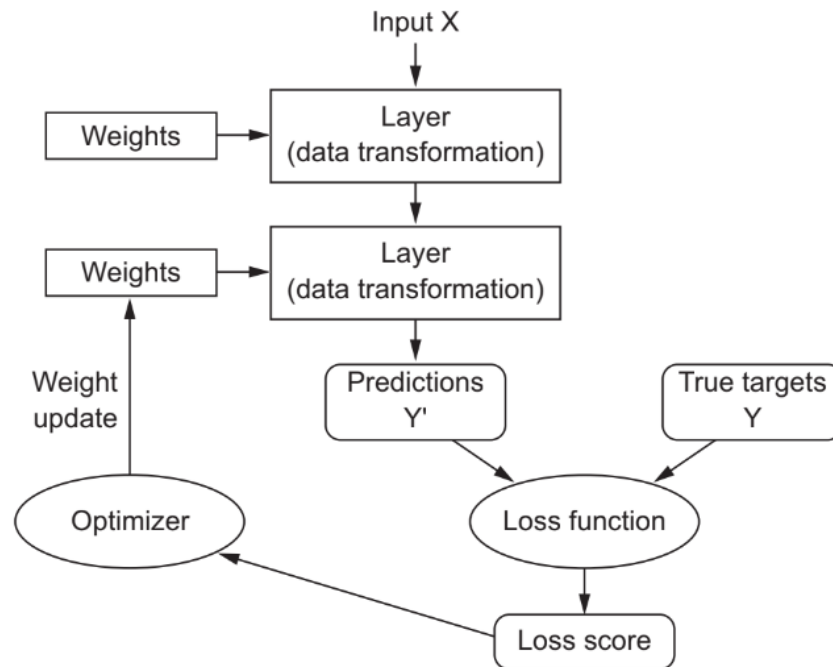
Este tipo de aprendizado é o caso mais comum. Consiste em aprender a mapear dados de entrada para os destinos conhecidos, que também chamados de anotações ou de rótulos (CHOLLET, 2018). Ou seja, de acordo com Marsland (2015), no processo de treinamento, um conjunto de dados de treino com as suas respectivas anotações é fornecido e, com base nesses dados, o algoritmo aprende e generaliza para responder corretamente a todas as entradas possíveis. No caso do presente trabalho, as imagens de satélite serão as entradas e as classes sociais os alvos.

## 5.4 APRENDIZADO PROFUNDO

O Aprendizado Profundo ("*Deep Learning*") é um subcampo específico do *Machine Learning* que utiliza estruturas matemáticas para aprender padrões. É uma abordagem sobre o aprendizado de características a partir de dados que enfatiza o aprendizado com camadas sucessivas de representações cada vez mais significativas. A palavra **Profundo** ou **Deep** em representa a ideia de várias camadas sucessivas e o seu número contribui para a profundidade do modelo. Essas representações são aprendidas por meio de modelos chamados redes neurais, que são estruturados em camadas literais empilhadas umas sobre as outras (CHOLLET, 2018). Na figura 1 é analisado o processo de treinamento das redes neurais profundas para os casos de aprendizado supervisionado.



Figura 1 – Representação do Deep Learning



Fonte: Chollet (2018)

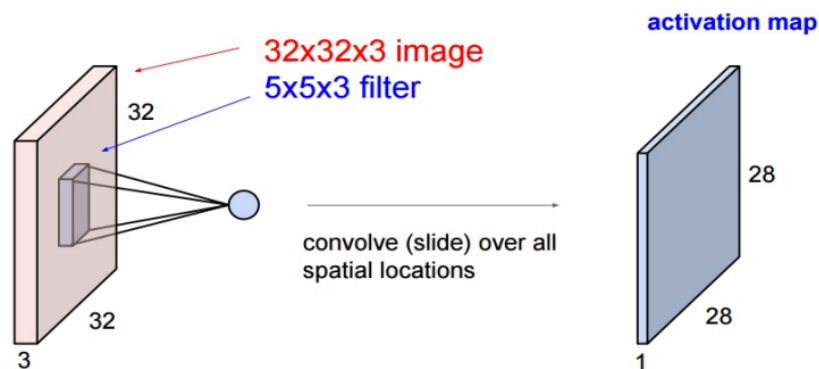
Os dados de entrada  $X$  passam pela primeira camada ("layer") e é multiplicada pelos pesos, que são parâmetros de uma camada, logo, nesse contexto, aprender significa encontrar um conjunto de valores para os pesos de todas as camadas da rede, de modo que a rede mapeie as entradas com as suas devidas saídas. Para controlar a saída de uma rede neural, é necessário calcular o erro da saída do modelo com o valor que era esperado através da função de perda ("Loss Function") da rede. Com os valores dos erros, é possível utilizá-los como um *feedback* para ajustar os valores dos pesos das camadas com o objetivo de diminuir a perda. Esse ajuste é o trabalho do otimizador ("Optimizer"). Inicialmente, os pesos da rede são atribuídos de forma aleatório ou pseudoaleatória, com isso os valores das perdas são altas, mas a cada exemplo que a rede processa, os pesos vão se ajustando e a perda vai diminuindo. Contudo, uma rede com perda mínima é aquela que a saída do modelo está mais próxima da saída esperada (CHOLLET, 2018).

#### 5.4.1 REDES NEURAS CONVOLUCIONAIS

As Redes Neurais Convolucionais ("Convolutional Neural Networks" - CNNs) é um tipo de rede neural que foi projetada para atuar com dados que possuem muitas dimensões, como imagens e vídeo (ZAFAR et al., 2018; KHAN et al., 2018). A ideia das CNNs foi popularizado através do trabalho dos LeCun et al. (1998), onde os autores introduziram a rede *LeNet-5* para o reconhecimento de dígitos escritos à mão.

As CNNs operam de forma semelhante às redes neurais padrões, porém utilizam da operação chamada **convolução** para filtrar em suas camadas as características úteis de suas entradas formando mapas de características, onde que apresenta os contornos horizontais e verticais necessários da imagem (ZAFAR et al., 2018; KHAN et al., 2018). Na Figura 2 é apresentado uma matriz da imagem de entrada com a dimensão  $32 \times 32 \times 3$  e um filtro ("kernel") de  $5 \times 5 \times 3$ , onde que é realizado a operação de convolução do filtro no decorrer de toda a imagem de entrada e tem como resultado uma matriz com dimensão  $28 \times 18 \times 1$ .

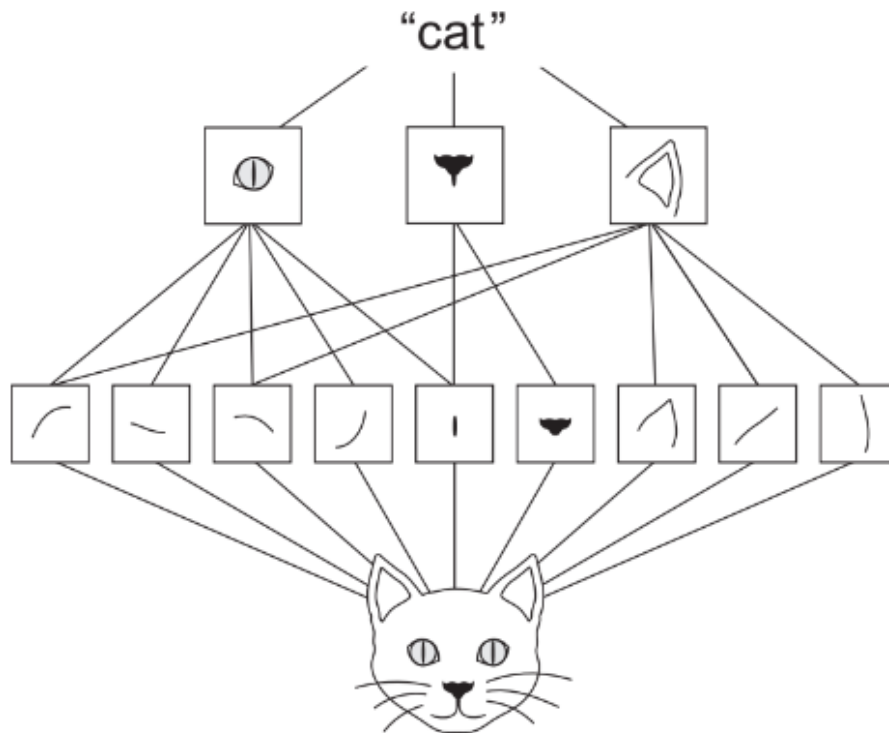
Figura 2 – Diagrama com a operação de convolução 2-D



Fonte: Zafar et al. (2018)

Uma das propriedades interessantes das redes convolucionais é que as mesmas conseguem aprender hierarquias espaciais de padrões, como na figura 3. Por exemplo, como visto na figura 3, a primeira camada de convolução aprenderá padrões locais e a segunda camada de convolução padrões maiores feitos pela primeira camada, e assim por diante. Isso permite que as CNNs aprendam com eficiência as configurações visuais complexas e abstratas.

Figura 3 – Exemplo de aprendizado das CNNs



Fonte: Chollet (2018)

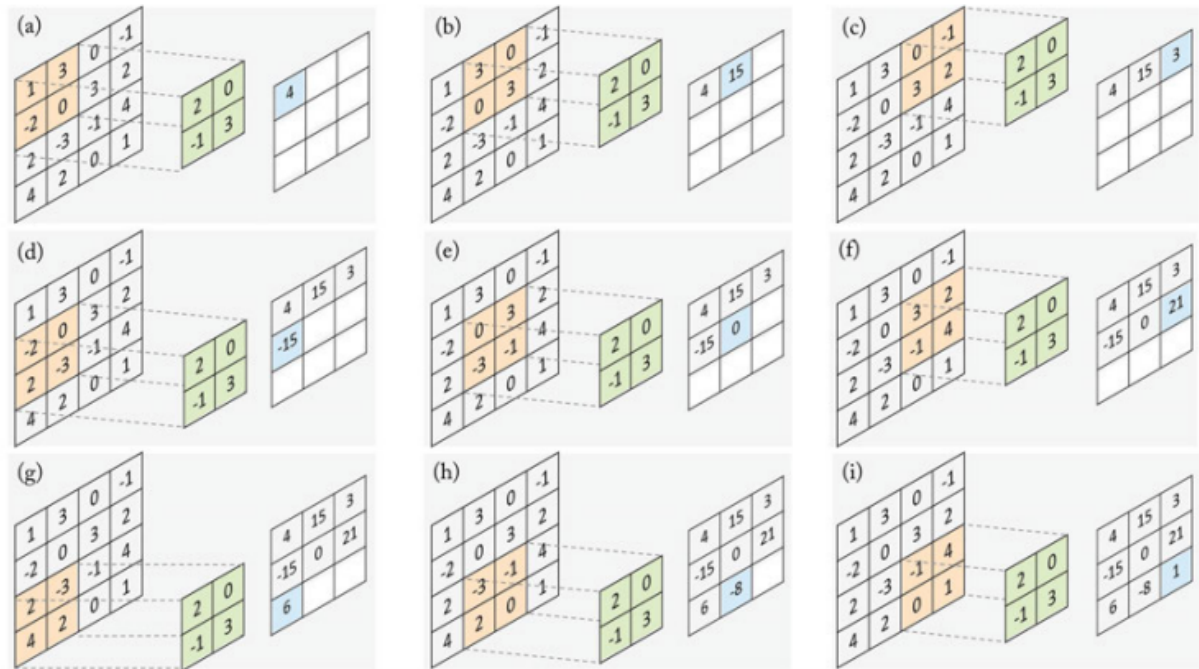
### • OPERAÇÃO DE CONVOLUÇÃO

De acordo com Zafar et al. (2018), a operação de convolução é uma operação linear que um sinal é resultado da soma do produto de dois outros sinais, é representada matematicamente como:

$$f(x, y) * g(x, y) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} f(n, m)g(x - n, y - m) \quad (2)$$

Na Figura 4 é ilustrado um exemplo da operação de uma camada de convolução. De (a) até (i) da Figura, é mostrado os cálculos realizados em cada etapa, pois o filtro é deslizado na imagem de entrada para calcular o valor correspondente no mapa de característica de saída. O filtro 2x2 é multiplicado com a mesma região de tamanho dentro da imagem de entrada 4x4 e os valores resultantes são somados para obter a matriz de saída (KHAN et al., 2018).

Figura 4 – Operação de Convolução em Imagens



Fonte: Khan et al. (2018)

Portanto, a diferença fundamental entre uma camada densamente conectada e uma camada de convolução é que as camadas densas (camadas das redes neurais padrões) aprendem características globais em seu dados de entrada, enquanto as camadas de convolução aprendem padrões locais, no caso de imagens, padrões encontrados em pequenas janelas 2D das entradas (CHOLLET, 2018).

#### 5.4.2 MODELOS DE CLASSIFICAÇÃO DE IMAGENS

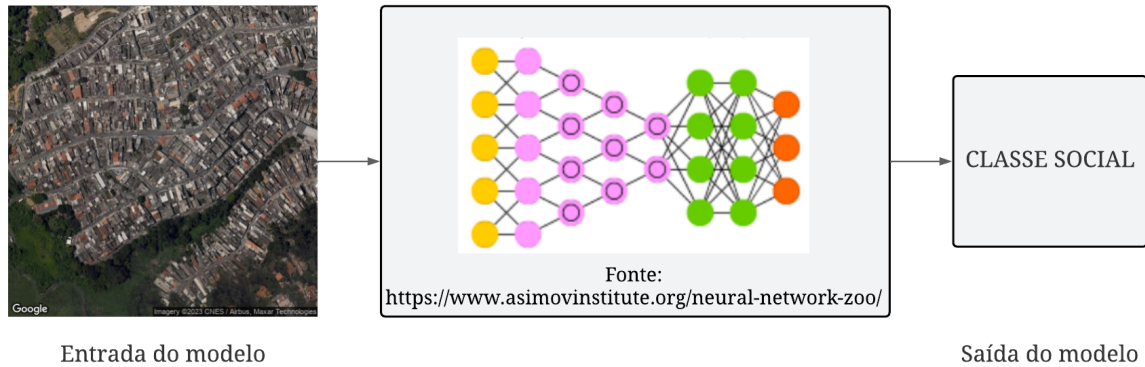
A classificação é uma das áreas mais importantes de Visão Computacional e reconhecimento de padrões. A função do classificador é usar um vetor para atribuir uma imagem ou região de interesse (RoI) a uma categoria (KHAN et al., 2018).

O grau da dificuldade depende da variação nos valores das características das imagens da mesma categoria em relação à diferença entre os valores das características das imagens de categorias diferentes. Porém, essa tarefa é bastante desafiadora devido aos problemas como presença de ruídos nas imagens (sombras, distorções, entre outras), *outliers*, ambiguidade, falta de rótulos e poucos dados para o treinamento (KHAN et al., 2018). Contudo, quanto melhor a qualidade da imagem e o rótulo é o mais adequado e o mais próximo da realidade, melhor é o resultado do modelo de classificação.

A figura 5 mostra o objetivo de um classificador de imagens na aplicação deste trabalho, que é justamente ter como entrada do modelo uma imagem de satélite e, a partir dos padrões aprendidos por ele na etapa de treinamento, obter uma categoria na

saída do modelo, que no caso desta aplicação, é a classe social.

Figura 5 – Objetivo Geral de modelos de classificação



Fonte: Própria

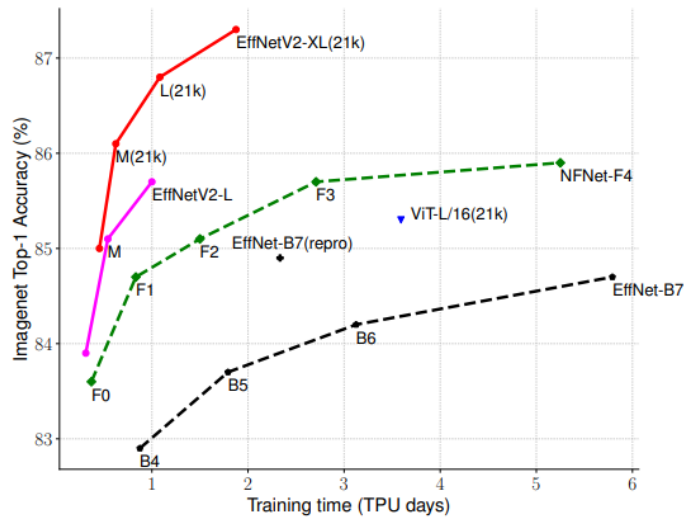
Para o desenvolvimento do projeto, será implementado a arquitetura de rede neural convolucional chamada *EfficientNetV2*, na qual é muito utilizada no problema de classificação de imagens.

- **EFFICIENTNET**

O *EfficientNet* é um dos modelos mais recente utilizado para a classificação de imagens na literatura, foi desenvolvido e publicado pelos pesquisadores da *Google* em 2021. Analisando o gráfico da Figura 6, o mesmo possui uma das melhores acurácias com o menor número de parâmetros. Esses resultados foram obtidos através do *ImageNet*<sup>2</sup>.

<sup>2</sup>Para mais detalhes, acessar: <https://www.image-net.org/>

Figura 6 – Tempo de Treino vs Acurácia



Fonte: Tan e Le (2021)

O *EfficientNetV2* é uma segunda família de arquiteturas de rede neurais profundas do estado da arte proveniente do trabalho de (TAN; LE, 2019), que apresenta o modelo com a finalidade de estudar sistematicamente o seu dimensionamento. A primeira família, cujos nomes vão de *EfficientNet-B0* à *EfficientNet-B7*, é resultado de um estudo que possibilita equilibrar, de forma cuidadosamente, a profundidade, largura e a resolução da rede para obter um melhor desempenho. De acordo com (TAN; LE, 2021), as diferenças entre a primeira e a segunda família são:

- A segunda família utiliza tanto o bloco *Mobile Inverted Bottleneck MBConv* do trabalho de (HOWARD et al., 2017) quanto o *fused-MBConv* do trabalho de Gupta e Tan (2019).
- EfficientNetV2* prefere taxa de expansão menor para *MBConv* já que taxas de expansão menores tendem a ter menos sobrecarga de acesso à memória.
- A segunda família prefere tamanhos menores de *kernel 3x3*, mas adiciona mais camadas para compensar o campo receptivo reduzido resultante da tamanho do *kernel* menor.
- EfficientNetV2* completamente remove o último estágio *stride-1* no *EfficientNet* original, talvez devido ao seu grande tamanho de parâmetro e acesso à memória a sobrecarga.

### 5.4.3 TRANSFER LEARNING

Para o treino dos modelos, será utilizado uma técnica chamada *Transfer Learning*, um método muito utilizado na literatura com a finalidade de treinar os modelos em um problema específico a partir das arquiteturas pré-treinadas em outro conjunto de dados.

Portanto, será implementado um rede neural pré-treinada em um grande conjunto de dados para solucionar o problema proposto. Esta técnica será utilizada devido à base de dados do projeto não ser muito grande, porque coletar os dados suficientes para treinar uma rede neural do zero é custoso e demanda tempo tanto para coletar quando para rotular (WEISS; KHOSHGOFTAAR; WANG, 2016; ZHUANG et al., 2019). Logo, os modelos pré-treinados serão implementados tendo a base de dados do *ImageNet* como ponto inicial.

#### 5.4.4 MÉTRICAS DE VALIDAÇÃO DOS MODELOS

Nesta subseção, é abordada sobre as métricas muito utilizadas para problemas de classificação que serão utilizadas para a validação dos modelos no projeto proposto. Essas métricas são extraídas da matriz de confusão, e são elas: Acurácia, Precisão (P), Revocação (R) e F1-Score.

- **MATRIZ DE CONFUSÃO**

A matriz de confusão é um dos métodos mais adequado para validar os algoritmos com os problemas de classificação (MARSLAND, 2015). Esse método consiste em uma matriz quadrada com os valores preditos pelo modelo e com os valores reais, portanto, é utilizada para analisar a quantidade de classificações corretas e incorretas do modelo.

Tabela 1 – Matriz de Confusão

Real	Previsto	
	Positivo	Negativo
Positivo	TP	FN
Negativo	FP	TN

Na Tabela 1, é representada uma matriz de confusão, onde que:

- Verdadeiro Positivo (*"True Positive"* - TP): é a frequência que o modelo previu a condição como positiva e era realmente positiva;
- Falso Positivo (*"False Positive"* - FP): é a frequência que o modelo previu a condição como positiva e, na verdade, era negativa;
- Verdadeiro Negativo (*"True Negative"* - TN): é a quantidade de vezes que o modelo previu a condição como negativa e realmente era negativa;
- Falso Negativo (*"False Negative"* - FN): é a frequência que o modelo previu a condição como negativa e, na verdade, era positiva.

Com a matriz de confusão, é possível calcular a Acurácia, Precisão (P), Revocação (R) e F1-Score, que são explicadas nos seguintes tópicos.

### • ACURÁCIA

A acurácia ("*accuracy*") é a métrica utilizada para analisar a porcentagem que o modelo acertou nas previsões possíveis. Portanto, é a divisão da soma das previsões corretas (Verdadeiros Positivos e Verdadeiros Negativos) com o somatório de todas as previsões (MARS LAND, 2015). Matematicamente, é definido como:

$$\text{Acurácia} = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

### • PRECISÃO

De acordo com Marsland (2015), a precisão ("*precision*") é a métrica utilizada para medir a porcentagem de verdadeiros positivos sobre o total que o modelo previu como positivo (Verdadeiros Positivos e Falsos Positivos). Por exemplo no problema deste trabalho, a precisão para a classe A será o número de imagens que o modelo afirmou como classe A e, que de fato é a classe A, dividido por todas as imagens que ele afirmou ser da classe A. Matematicamente, é definido como:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (4)$$

### • REVOCAÇÃO

A revocação ("*recall*") é a métrica utilizada para analisar a proporção de positivos identificados corretamente. Portanto, é a razão entre os verdadeiros positivos sobre a soma de verdadeiros positivos com falsos negativos (MARS LAND, 2015). No caso do presente trabalho, a revocação para a classe A será o número de imagens que o modelo afirmou como classe A e, que de fato é a classe A, dividido pelo número total das imagens da classe A. Matematicamente, é definido como:

$$\text{Revocação} = \frac{TP}{TP + FN} \quad (5)$$

### • F1-SCORE

De acordo com Marsland (2015), o F1-Score é uma média harmônica entre a precisão e a revocação que é utilizada como relação e uma análise geral da precisão e revocação:

$$\text{F1-score} = 2 * \frac{\text{precisão} * \text{revocação}}{\text{precisão} + \text{revocação}} \quad (6)$$



## 5.5 PYTHON

*Python* é uma linguagem de programação de alto nível e de código aberto. É uma linguagem otimizada a fim de ter uma qualidade de software, produtividade do desenvolvedor, portabilidade do programa e integração de componentes. O seu recurso mais importante é o suporte para vários paradigmas da programação, como orientação a objetos e programação imperativa e funcional, além de suportar sistemas dinâmicos e obter um gerenciamento automático de memória, também possui uma grande biblioteca padrão abrangente (LUTZ, 2010; SRINATH, 2017). É uma linguagem interpretada, porém, seus intérpretes estão disponíveis para muitos sistemas operacionais (SRINATH, 2017).

## 5.6 OPENCV

OpenCV (*Open Source Computer Vision Library*- Figura 7) é uma biblioteca de software para visão computacional e *machine learning* de código aberto. Foi desenvolvido com a finalidade de fornecer aos aplicativos de visão computacional uma infraestrutura comum, além de acelerar o uso da percepção da máquina em produtos comerciais. Essa biblioteca possui mais de 2.500 algoritmos otimizados, dentre eles, estão os algoritmos de visão computacional e de aprendizado de máquina, tanto os clássicos quanto os mais recentes. O OpenCV obtém interfaces para linguagens C++, Python, Java e MATLAB, porém, é escrito nativamente em C++. O mesmo suporta os sistemas operacionais *Windows, Linux, Android e Mac OS* (OPENCV, 2022).

Figura 7 – OpenCV Logo



Fonte: OpenCV (2022)

## 5.7 TENSORFLOW E KERAS

O *TensorFlow* (Figura 8) é uma das plataformas completas mais utilizada de código aberto para *Machine Learning* e *Deep Learning*, onde que os usuários iniciantes e avançados podem criar e implementar os modelos para computadores, dispositivos móveis, Web e nuvem. Oferece várias ferramentas de dados que auxiliam na consolidação, limpeza e pré-processamento de dados em grande escala. Além disso, apoia o treinamento

distribuído, a iteração imediata de modelos e a depuração com o *Keras*. Apresenta ferramentas como *Model Analysis* e *TensorBoard* para o monitoramento do ciclo de vida do modelo (TENSORFLOW, 2022).

Figura 8 – TensorFlow Logo



Fonte: TensorFlow (2022)

O *TensorFlow* disponibiliza recursos para a implementação dos modelos em qualquer ambiente, como servidores, dispositivos de borda, navegadores, dispositivos móveis, microcontroladores, CPUs (*Central Processing Units*), GPUs (*Graphics Processing Units*) e FPGAs (*Field Programmable Gate Array*). Apresenta o *TensorFlow Serving* para executar modelos de *machine learning* em escala de produção nos processadores mais avançados atualmente, como as TPUs (*Tensor Processing Unit*) desenvolvida pela *Google*. O *TensorFlow Lite* é um *framework* utilizado para a execução dos modelos em dispositivos móveis, dispositivos de computação de borda e em microcontroladores, enquanto que o *framework TensorFlow.js* permite a execução dos modelos de *machine learning* para os navegadores da Web (TENSORFLOW, 2022).

O *Keras* (Figura 9) é uma API (*Application Programming Interface*) de alto nível programada em Python e executada por cima do *TensorFlow 2*. O seu desenvolvimento foi com a finalidade de permitir uma experimentação rápida reduzindo a carga cognitiva do desenvolvedor para liberá-lo a se concentrar nas partes do problema que é essencial. De acordo com *Keras* (2022), essa API segue o princípio de "Ser capaz de ir da ideia ao resultado o mais rápido possível é fundamental para fazer uma boa pesquisa". Possui uma interface acessível e produtiva para resolver problemas de *Deep learning*, fornece blocos de construção para desenvolver e enviar soluções com alta velocidade de iteração. É possível executar o *Keras* em TPU ou em grandes *clusters* de GPUs e exportar os modelos para execução nos navegadores ou dispositivos móveis (KERAS, 2022).

Figura 9 – Keras Logo



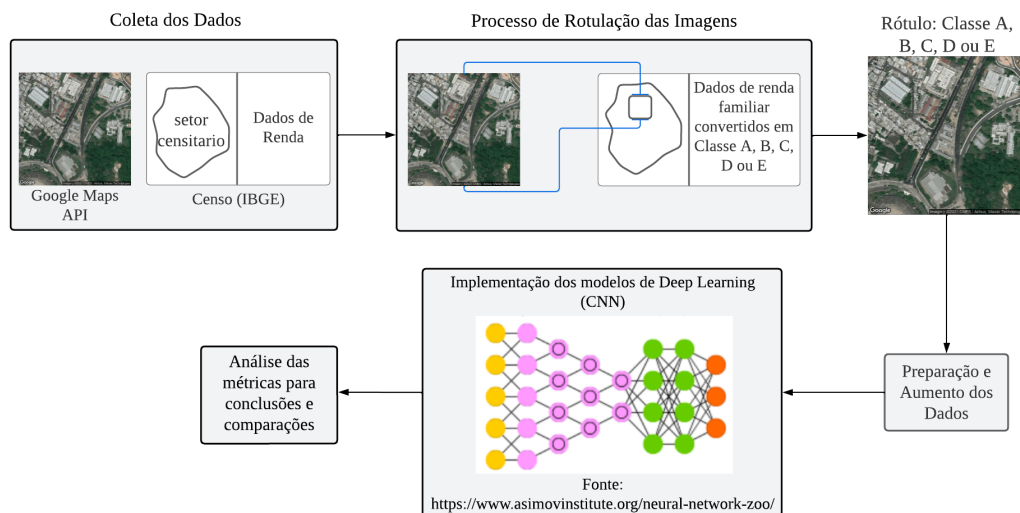
Fonte: Keras (2022)

## 6 METODOLOGIA

Esse estudo teve por finalidade realizar uma pesquisa de natureza aplicada. Para alcançar os objetivos propostos, foi utilizada uma abordagem quali-quantitativa. Com o intuito de conhecer a problemática sobre a área de estudo, foi realizada uma pesquisa exploratória e bibliográfica (GIL, 2019). A seguir são descritos os procedimentos metodológicos aplicados, bem como os materiais utilizados.

A implementação de modelos de Visão Computacional para o problema de classificação de classes sociais das cidades de Manaus (AM) e São Paulo (SP) foi dividido em 4 etapas: criação da base de dados ("dataset"), preparação dos dados, implementação da arquitetura de CNN e, por último, análise dos resultados para avaliar a eficácia dos modelos obtidos. Na Figura 10 é apresentado o diagrama com os processos e as etapas que foram realizadas no desenvolvimento do trabalho.

Figura 10 – Diagrama das Etapas do Desenvolvimento do Trabalho



Fonte: Própria

### 6.1 CRIAÇÃO DO CONJUNTO DE DADOS

Para obter um bom resultado nos problemas com modelos de classificação, é necessário um conjunto de dados bem estruturados e condizentes com a realidade, no caso de dados de imagens, os rótulos ou *labels* das imagens precisam estar adequados com o que está contido nas mesmas. Logo, o processo da criação do *dataset* foi dividido em 3 etapas que serão abordados a seguir:

- Coleta das imagens de Satélite através da API do *Google Static Maps*<sup>3</sup> das cidades de Manaus (AM) e São Paulo (SP);

<sup>3</sup><https://developers.google.com/maps/documentation/maps-static/overview>

- b) Coleta e análise das tabelas de Renda do Censo 2010 visto que os dados do Censo de 2022 não estão disponibilizadas no presente momento;
- c) Processo de rotulação das imagens com os dados de renda analisados na etapa b.

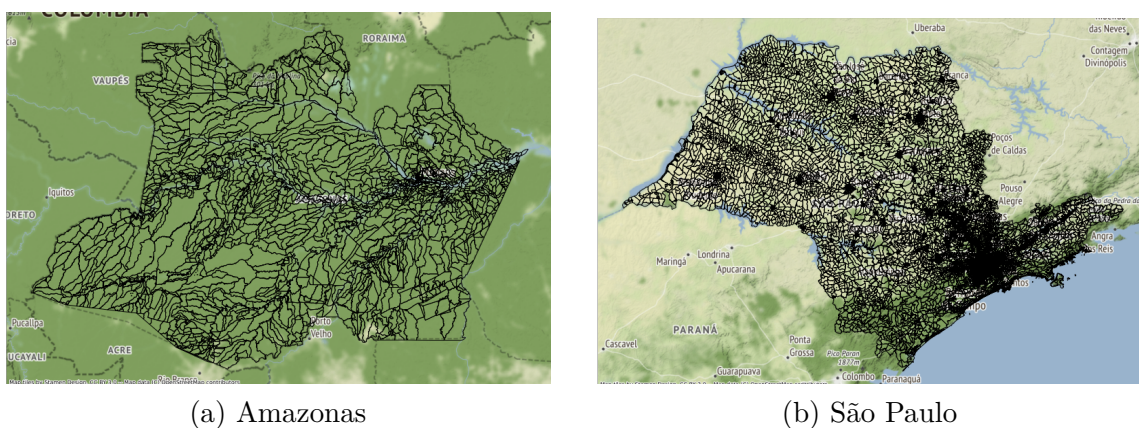
### 6.1.1 COLETA DAS IMAGENS DE SATÉLITES

O primeiro passo do processo de coleta das imagens é utilizar as malhas dos setores censitários do Censo Demográfico de 2010 das cidades analisadas. Essas malhas são disponibilizadas em formas de dados geográficos em um arquivo do tipo *shapefile*, um formato que armazena informações georreferenciadas, ou seja, fornece localizações geométricas e atributos de características geográficas como pontos, linhas e polígonos, para cada estado nacional, foram utilizadas as malhas do estado do Amazonas<sup>4</sup> e do estado de São Paulo<sup>5</sup>. Em ambos os arquivos, são encontradas as devidas informações relevantes para o desenvolvimento do trabalho:

- a) O código de identificação do setor censitário;
- b) O tipo de cada setor: rural ou urbano;
- c) O nome do município em que o setor se localiza;
- d) A geometria do polígono de cada setor censitário.

A leitura e toda a análise dos dados geográficos presente nos arquivos em *shapefile* foram utilizando as funções da biblioteca GeoPandas. Nas figuras 14a e 14b são visualizadas os setores censitários dos estados do Amazonas e de São Paulo, respectivamente.

Figura 11 – Setores censitários de cada estado



Fonte: Própria

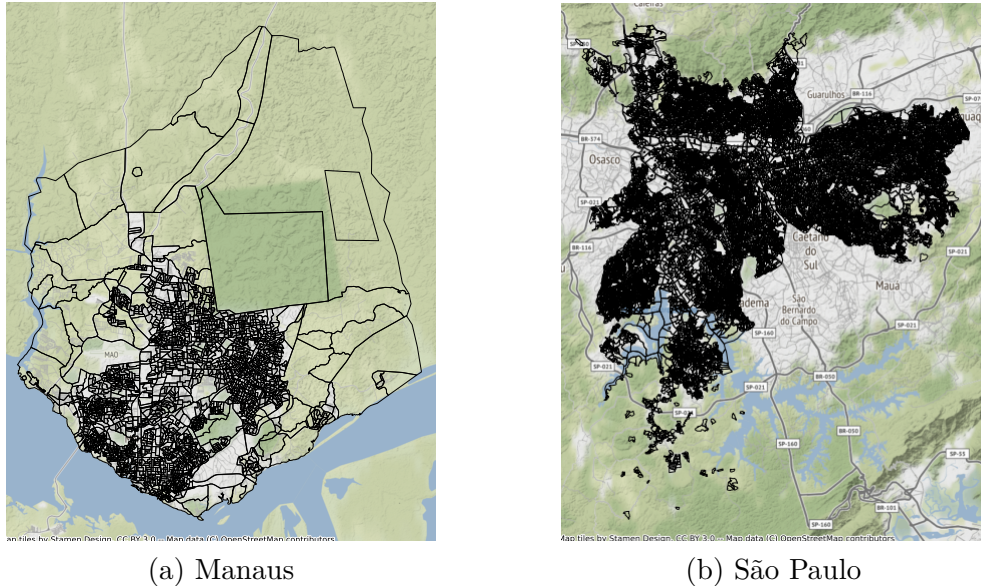
<sup>4</sup>Para realizar o *download* da malha do Amazonas: ([https://geoftp.ibge.gov.br/organizacao\\_do\\_territorio/malhas\\_territoriais/malhas\\_de\\_setores\\_censitarios\\_divisoes\\_intramunicipais/censo\\_2010/setores\\_censitarios\\_shp/am/am\\_setores\\_censitarios.zip](https://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios_divisoes_intramunicipais/censo_2010/setores_censitarios_shp/am/am_setores_censitarios.zip))

<sup>5</sup>Para realizar o *download* da malha de São Paulo: ([https://geoftp.ibge.gov.br/organizacao\\_do\\_territorio/malhas\\_territoriais/malhas\\_de\\_setores\\_censitarios\\_divisoes\\_intramunicipais/censo\\_2010/setores\\_censitarios\\_shp/sp/sp\\_setores\\_censitarios.zip](https://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios_divisoes_intramunicipais/censo_2010/setores_censitarios_shp/sp/sp_setores_censitarios.zip))



Com as informações do município e do tipo da área de cada setor censitário, foi possível filtrar para obter somente os setores censitários das áreas urbanas dos municípios de Manaus (AM) e São Paulo (SP). Os setores de cada cidade, na qual é a área urbana de cada município, são apresentados na figura 12.

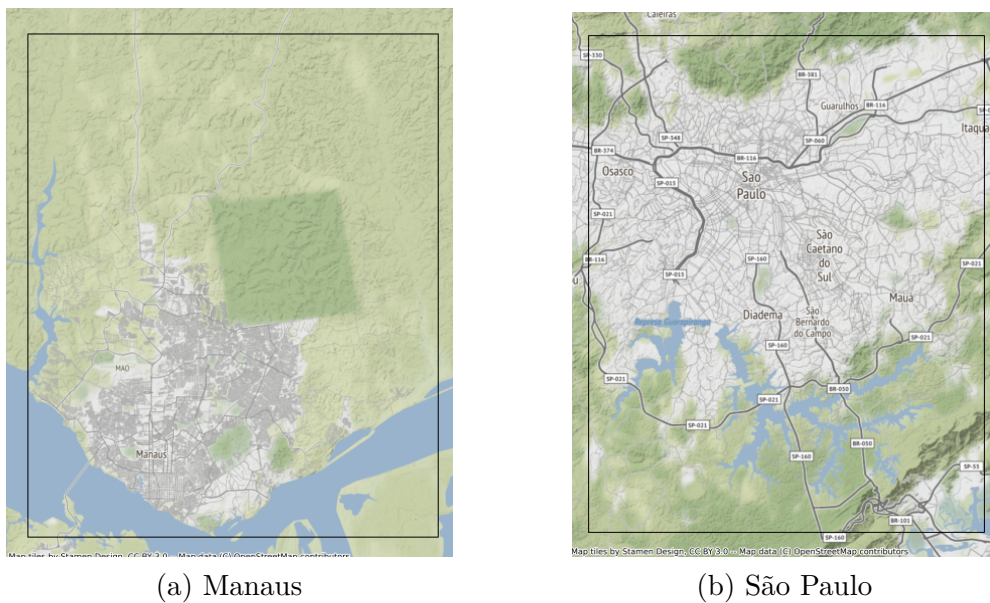
Figura 12 – Setores censitários de cada cidade



Fonte: Própria

Através dos polígonos dos setores censitários, foi obtido as coordenadas máximas e mínimas dos limites de cada cidade e, com isso, foi feito um polígono no formato de um retângulo com as coordenadas como apresentados na Figura 13.

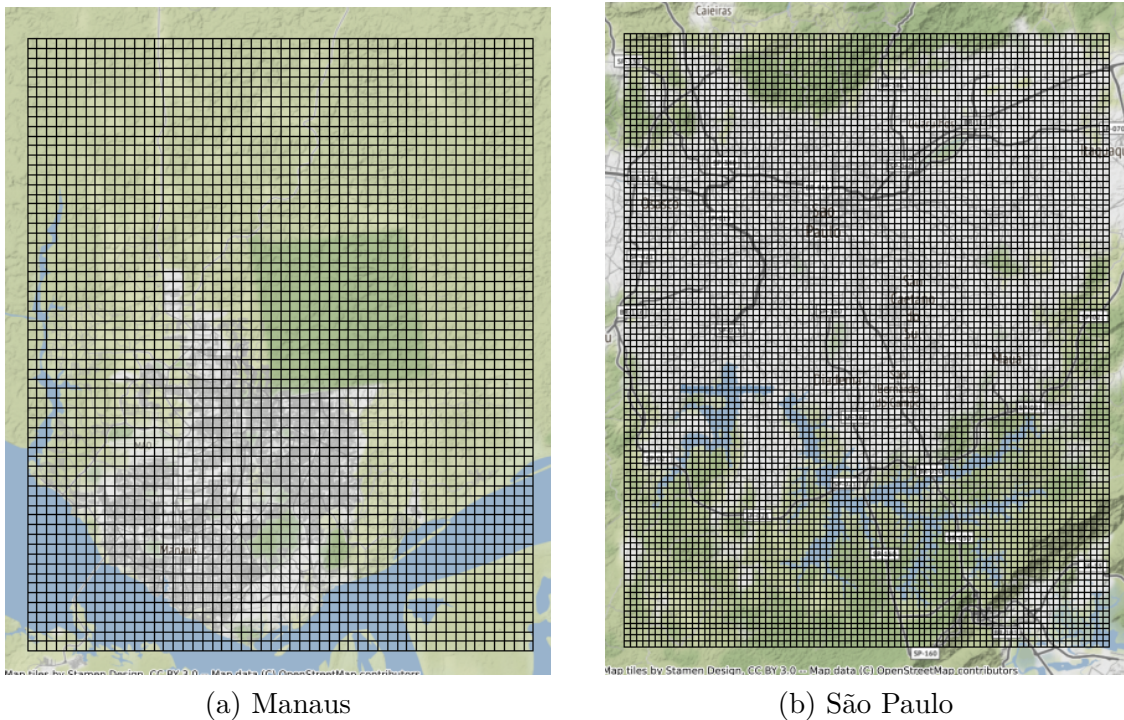
Figura 13 – Limites máximo e mínimo de cada cidade no formato de um retângulo



Fonte: Própria

A partir disso, foi analisado nos trabalhos relacionados a área do espaço geográfico das imagens coletadas para obter como ponto de referência. O trabalho de (CASTRO; ÁLVAREZ, 2022), utiliza uma faixa de 740m à 810m como resolução espacial. Contudo, foi optado por 650 x 650 metros o máximo de área que as imagens iriam representar e, assim, dividiu-se os retângulos da Figura 13 em *grids* de acordo com resolução espacial definida.

Figura 14 – Total de *Grids* envolvendo todo o retângulo de cada cidade

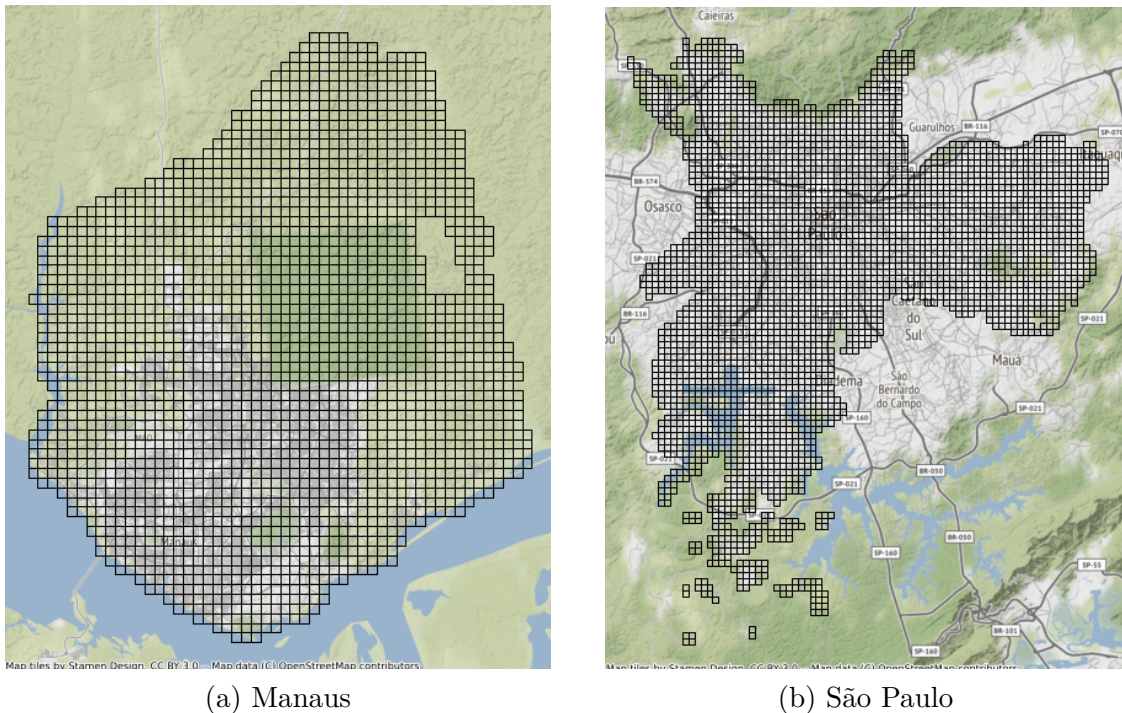


Fonte: Própria

O próximo passo foi filtrar os *grids* da Figura 14 para obter somente os que fazem interseção com os setores censitários da Figura 12 resultando nos *grids* ilustrados na Figura 15.



Figura 15 – *Grids* dentro dos limites dos setores censitários em cada cidade



Fonte: Própria

A partir dos polígonos da Figura 15, calculou-se os centroides, centro geométrico, de cada um dos *grids* apresentados. Esses centroides são os pontos com as coordenadas geográficas que será o centro de cada imagem de satélite coletada, visto que, para realizar uma requisição pela a API do Google Maps Static, é necessário os seguintes parâmetros:

- a) Chave API ("*API Key*"): uma chave criada na plataforma de nuvem do Google para utilizar a devida API;
- b) Centro ("*center*"): a coordenada geográfica do centro da imagem;
- c) O nível do Zoom da imagem;
- d) O tamanho que a imagem coletada tem.

Os parâmetros do zoom e do tamanho da imagem servem para ajustar a resolução espacial das imagens que serão coletadas. A escolha do nível do zoom foi de 17 e o tamanho da imagem foi de 480 x 480 pixels. Contudo, a área geográfica é de, aproximadamente, 575 x 575 metros visto que, para o zoom de 17, a relação entre metros e pixels é de 1.194 m/pixels. Com essas informações as imagens foram coletadas para ambas cidades e, como o Google Maps Static API coleta as imagens de satélite atuais das cidades, as imagens do conjunto de dados foram do ano 2022/2023. Cada imagem possui o seu *id* correspondendo ao dos *grids* para permanecer uma relação entre a imagem salva com o seu respectivo *grid*. Exemplo das imagens coletadas para cada cidade é apresentada na Figura 16.

Figura 16 – Exemplo das imagens coletadas de ambas as cidades



Fonte: Própria

### 6.1.2 DADOS DE RENDA DO CENSO DE 2010

Como já mencionado, as imagens de satélites foram rotuladas ou anotadas de acordo com os dados de renda do Censo Demográfico de 2010. O Censo disponibiliza 3 principais tabelas para cada estado nacional com informações completas sobre a renda em função dos setores censitários, sendo elas: Domicílio Renda, Pessoa Renda e Responsável Renda.

- **Domicílio Renda:** as informações apresentadas nesta tabela são referentes ao total do rendimento nos domicílios em todo o setor censitário. Analisando as informações, foi optado por não utilizar esta tabela para rotular as imagens por possuir somente o número total de rendimento mensal dos domicílios e o número de domicílios para cada classe social e, com isso, poderia obter uma faixa de erro no cálculo para determinar qual classe (classe A, B, C, D e E) cada setor censitário pertenceria, visto que, não tem a informação de uma renda média mensal dos domicílios para cada parte do salário mínimo.
- **Pessoa Renda:** esta tabela do censo apresenta diversas informações, são em torno de 134 campos ou colunas. Porém, foram selecionadas as informações apresentadas na Tabela 2 que são as que possuem a coerência para o desenvolvimento do trabalho. Os dados contidos são referentes ao número de pessoas com 10 anos ou mais que colaboram com a renda em cada setor censitário.
- **Responsável Renda:** já esta tabela também apresenta os mesmo dados contidos na da Pessoa Renda, portanto, foram selecionadas os campos apresentados na Tabela 3. Porém, as estatísticas são referentes ao responsável do domicílio em cada setor



censitário.

Tabela 2 – Informações selecionadas na tabela de Pessoa Renda

Nome da Variável	Descrição da Variável
Cod setor	Código do setor censitário
V011	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de até $\frac{1}{2}$ salário mínimo
V012	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de $\frac{1}{2}$ a 1 salário mínimo
V013	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 1 a 2 salários mínimos
V014	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 2 a 3 salários mínimos
V015	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 3 a 5 salários mínimos
V016	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 5 a 10 salários mínimos
V017	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 10 a 15 salários mínimos
V018	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 15 a 20 salários mínimos
V019	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade com rendimento nominal mensal de mais de 20 salários mínimos
V022	Total do rendimento nominal mensal das pessoas de 10 anos ou mais de idade

Fonte: IBGE (2010)

Tabela 3 – Informações selecionadas na tabela de Responsável Renda

Nome da Variável	Descrição da Variável
Cod setor	Código do setor censitário
V011	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de até $\frac{1}{2}$ salário mínimo
V012	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de $\frac{1}{2}$ a 1 salário mínimo
V013	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 1 a 2 salários mínimos
V014	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 2 a 3 salários mínimos
V015	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 3 a 5 salários mínimos
V016	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 5 a 10 salários mínimos
V017	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 10 a 15 salários mínimos
V018	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 15 a 20 salários mínimos
V019	Total do rendimento nominal mensal das pessoas responsáveis com rendimento nominal mensal de mais de 20 salários mínimos
V022	Total do rendimento nominal mensal das pessoas responsáveis

Fonte: IBGE (2010)

Contudo, as tabelas de Pessoa Renda e de Responsável Renda foram utilizadas para o processo de rotulação das imagens de satélites. Para isso, primeiramente foi verificado quantos salários mínimos é a classe A, B, C, D e E.

- a) Classe A: renda superior à 20 salários mínimos;
- b) Classe B: renda de 10 à 20 salários mínimos;
- c) Classe C: renda de 4 à 10 salários mínimos;
- d) Classe D: renda de 2 à 4 salários mínimos;
- e) Classe E: renda menor que 2 salários mínimos.

Obtendo o número de salários mínimos para cada classe e com as informações das tabelas 2 e 3 é possível verificar a proporção que cada classe tem em cada setor censitário. Como as tabelas tem as mesmas estruturas de nomes das variáveis com as suas respectivas descrições, então, é mostrado a seguir a relação de cada campo das tabelas para calcular a proporção de cada classe social:

- a) Classe A:  $V019/V022$ ;
- b) Classe B:  $(V017 + V018)/V022$ ;
- c) Classe C:  $V016/V022$ ;
- d) Classe D:  $(V014 + V015)/V022$ ;
- e) Classe E:  $(V011 + V012 + V013)/V022$ .

A partir disso, é possível declarar a classe social do setor censitário sendo a classe com maior proporção. Por exemplo, em um setor que possui uma renda total sendo  $X$  reais, a classe A tem uma participação de 5% do total da renda, a classe B tem de 20%, a classe C tem 40%, a classe D tem 20% e a classe E tem 15%, neste caso, esse setor censitário seria da classe C visto que a sua proporção do total da renda seria maior do que das outras classes, ou seja, é a classe predominante nesse setor censitário.

Todavia, como foi utilizado as duas tabelas, Pessoa Renda e Responsável Renda, 1 setor censitário de cada estado tem 2 classes predominantes. Isso porque pela tabela da Pessoa Renda, um determinado setor pode ser de Classe D, porém, pela tabela do Responsável Renda pode ser de Classe E. Logo, no desenvolvimento do trabalho, foi aplicado o método para os 2 casos e, assim, compará-los e verificar nos resultados qual seria a melhor forma de anotar as imagens para o modelo.

### 6.1.3 PROCESSO DE ANOTAÇÃO DAS IMAGENS

A próxima etapa da construção do conjunto de dados para o treinamento é o processo de anotar ou rotular as imagens de satélites com as classes estabelecidas na

etapa anterior. Ou seja, é a integração das imagens com as classes de acordo com as tabelas da Pessoa Renda e do Responsável Renda.

Para isso, precisou-se utilizar novamente os polígonos dos setores censitários apresentados na Figura 12 e as coordenadas geográficas dos *grids* da Figura 15, como os *id* dos *grids* correspondem com as das imagens, no caso, o *id* das imagens estão nos nomes dos arquivos das imagens, logo, pode-se analisar somente a relação dos polígonos dos setores censitários e dos *grids*, que os seus rótulos serão os rótulos das imagens.

Primeiramente, foi listado os setores censitários que intersectavam cada *grid* como apresentado na Figura 17. Cada setor obteve a classe social na etapa anterior, então para definir qual classe o *grid* pertence, visto que, 1 setor pode ser de classe C e o outro de classe D, foi analisado de 2 formas diferentes:

1. Pela área de interseção de cada setor censitário com o *grid*, então neste caso, seria a classe do setor com a maior área dentro do *grid*, ou seja, pelo setor predominante. No exemplo seria a classe do setor de cor vermelha.
2. Pelo centro geométrico (centroide) do *grid*. Neste caso, seria a classe social do setor censitário que contém o ponto do centroide. No exemplo dado na Figura 17 também neste caso seria a classe do setor censitário de cor vermelha.

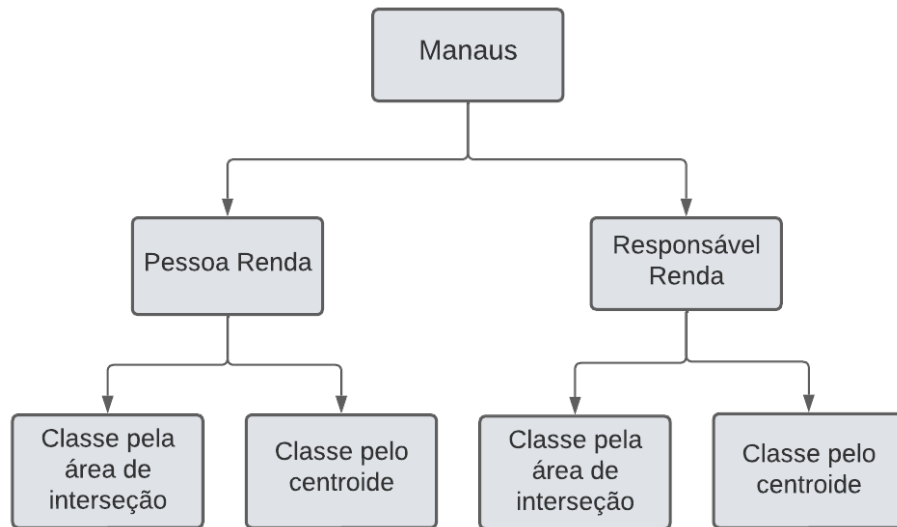
Figura 17 – Exemplo da Interseção de 1 *grid* com 2 setores censitários



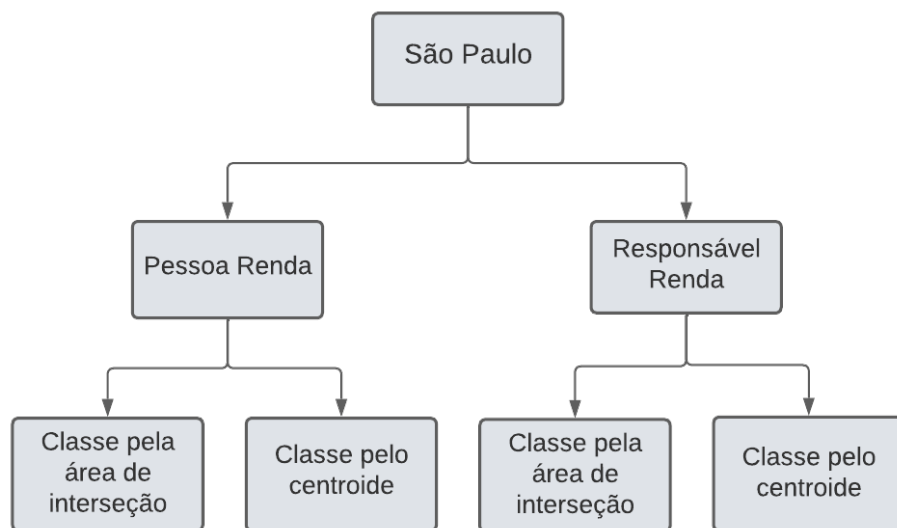
Fonte: Própria

Contudo, feito esses 2 métodos para todos os *grids* de Manaus (AM) e de São Paulo (SP), foi obtido 4 formas de anotar uma imagem de satélite para cada cidade, ou seja, 4 conjuntos de dados diferentes, o esquema é apresentado na Figura 18. Ou seja, uma única foto tem 4 métodos de definir a sua classe social sendo que um método pode informar que a imagem é de classe A e outro método dizer que é classe B. Portanto, foram treinados os modelos utilizando cada forma de anotar as imagens para poder analisar e comparar os resultados e, assim, definir qual método pode ser mais adequado.

Figura 18 – Diagrama para a rotulação das imagens para cada cidade



(a) Manaus



(b) São Paulo

Fonte: Própria

## 6.2 PREPARAÇÃO DO CONJUNTO DE DADOS E TREINAMENTO DOS MODELOS

Obtendo as imagens com seus respectivos rótulos, o próximo passo foi preparar o conjunto de dados e, assim, realizar os devidos treinamentos dos modelos. Nesta etapa, foi utilizado 2 abordagens diferentes:

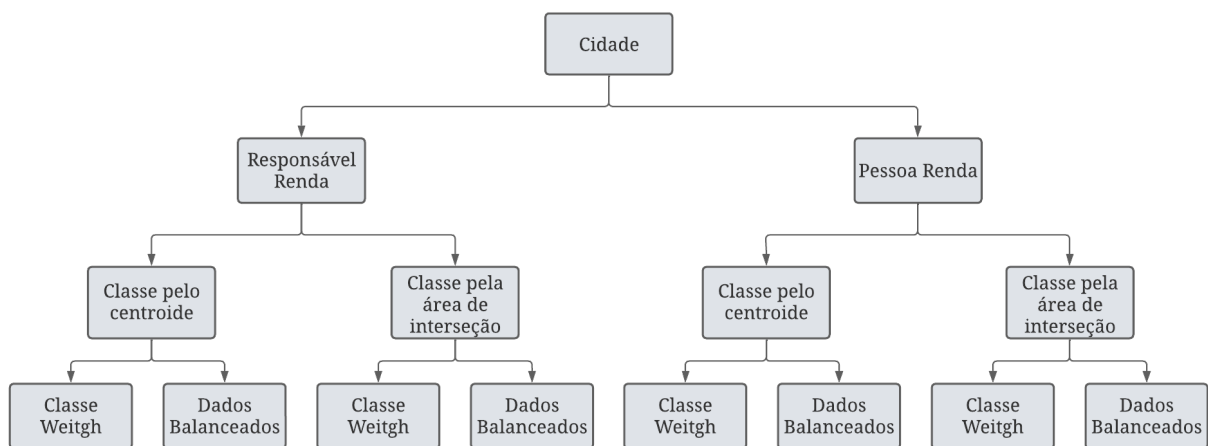
a) Treinamento com aumento de dados e com os pesos das classes (técnica chamada de

*class weight*);

b) Treinamento utilizando a base de dados balanceada.

Para cada método de rotular as imagens descritas na subseção anterior, foram aplicadas as 2 abordagens acima. Contudo, foram treinadas e testadas 16 modelos de redes neurais sendo 8 para cada cidade (Manaus e São Paulo) visto que são 4 formas de rotular as imagens. Todas essas abordagens foram utilizando a arquitetura do *EfficientNetV2* através da API do *Keras - Tensorflow* com a técnica do *transfer learning* (Seção 5.4.3) tendo os pesos iniciais sendo do *ImageNet*.

Figura 19 – Diagrama de Treinamento



Fonte: Própria

Em todos os casos, a base de dados foi dividida em dados de treino, dados de validação e dados de teste. Os dados de treino são utilizados para treinar o modelo, ou seja, são os dados que a rede neural aprende os padrões durante o treinamento, os dados de validação são os dados que a rede neural utiliza para testar durante o treinamento e a base de teste serve para validar e comprovar o aprendizado do modelo sendo utilizado depois do processo de treino.

Na tabela 4, apresenta a distribuição das classes sociais para todas os 4 conjuntos de dados formados através das técnicas de rotulação das imagens, tanto para a cidade de Manaus e São Paulo. Por exemplo, na cidade de Manaus, utilizando a tabela da Pessoa Renda do Censo com o método de verificar a classe social através do setor censitário que contém o centroide do *grid* da imagem apresenta 139 imagens de satélite para a classe A, 116 imagens de satélite para a classe B, 114 imagens para a classe C, 104 imagens para a classe D e 1378 imagens para a classe E obtendo no total 1851 imagens para este conjunto de dados.

Tabela 4 – Distribuição das classes sociais nos conjuntos de dados

Cidade	Tabela do Censo	Forma de Rotulação	A	B	C	D	E	Total
Manaus	Pessoa	Centroide	139	116	114	104	1378	1851
		Setor predominante	162	130	125	107	1459	1983
	Responsável	Centroide	137	117	139	159	1299	1851
		Setor predominante	163	133	152	161	1374	1983
São Paulo	Pessoa	Centroide	298	141	366	873	984	2662
		Setor predominante	316	142	399	989	1274	3120
	Responsável	Centroide	348	194	434	1006	679	2661
		Setor predominante	316	142	399	989	1274	3120

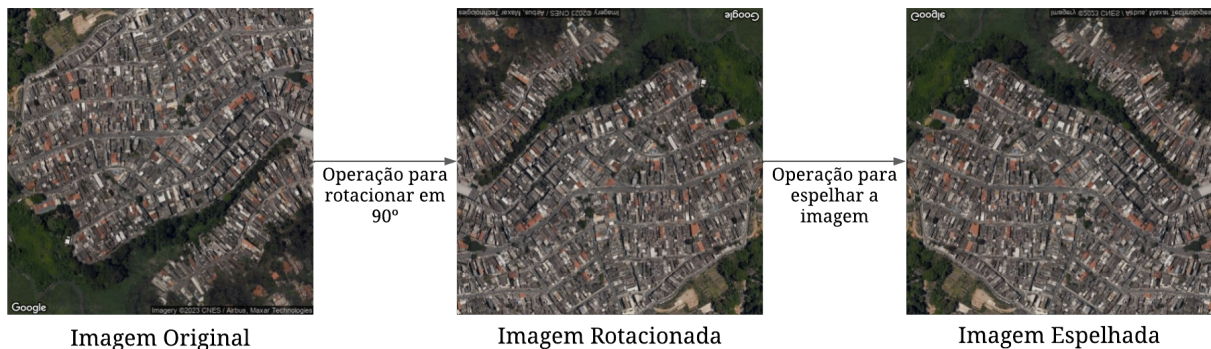
Fonte: Própria

### 6.2.1 PREPARAÇÃO DO CONJUNTO DE DADOS PARA TREINAMENTO COM O AUMENTO DE DADOS E PESOS DAS CLASSES

Primeiramente, foi dividido a base de dados de teste com a taxa de 20% de cada classe social a partir da base total e, dos 80% restante, 20% foi da base de validação. Em seguida, foi criada uma *pipeline* de transformações de imagens com o objetivo de realizar o aumento de dados utilizando a biblioteca *Albumentations*. A *pipeline* consistiu em:

- Rotacionar as imagens em 90°;
- Realizar o *flip* ("espelhamento") da imagem no sentido horizontal.

Na Figura 20 apresenta um exemplo da imagem de São Paulo com as transformações da *pipeline* realizada.

Figura 20 – *Pipeline* das transformações no aumento de dados de imagens

Fonte: Própria

O aumento de dados foi aplicado nas imagens de satélite na base de treino para somente as categorias minoritárias, que são as categorias que menos possuem imagens, geralmente, as classes sociais A e B como visto na tabela 4 com a distribuição das classes sociais nos conjuntos de dados totais. A tabelas 5 mostra a distribuição das categorias de acordo com cada conjunto de dados para o treino.

Tabela 5 – Distribuição das classes sociais nos conjuntos de dados de treino

Cidade	Tabela do Censo	Forma de Rotulação	A	B	C	D	E	Total
Manaus	Pessoa	Centroide	355	294	291	266	882	2088
		Setor predominante	309	249	240	204	934	1936
	Responsável	Centroide	262	223	266	305	831	1887
		Setor predominante	312	254	290	307	880	2043
São Paulo	Pessoa	Centroide	381	269	467	558	629	2304
		Setor predominante	403	271	510	633	815	2632
	Responsável	Centroide	445	372	555	643	434	2449
		Setor predominante	403	271	510	633	815	2632

Fonte: Própria

Além da técnica de aumento de dados, também foi aplicado os pesos das categorias no momento do treinamento. Logo, as categorias minoritárias obtiveram pesos maiores do que as categorias majoritárias. Por exemplo, em uma base de dados com 10 amostras para um problema binário, se a base for balanceada (50% por 50%), os pesos das classes são iguais, porém, se a base for 70% por 30%, a categoria que possui 3 amostras terá mais peso do que a que possui 7 amostras. Contudo, esta técnica é muito utilizada para problemas com o conjunto de dados desbalanceado, ou seja, uma classe tem mais amostras do que as outras. Então, para calcular os pesos, utilizou-se a equação 7:

$$class\_weight = \frac{n\_samples}{n\_classes * np.bincount(y)} \quad (7)$$

Onde que o  $n\_samples$  é o número de amostras de uma classe, o  $n\_classes$  é o número de classes e  $np.bincount(y)$  é um *numpy array* com o número de ocorrências de cada categoria.



### 6.2.2 PREPARAÇÃO DO CONJUNTO DE DADOS PARA TREINAMENTO COM A BASE BALANCEADA

Para realizar o treinamento com a base de dados balanceada, foi utilizada a técnica do *undersampling*, no qual é uma técnica que consiste em manter todos os dados da classe com menor frequência (classe minoritária) e diminuir a quantidade dos que estão na classe de maior frequência (classe majoritária). Na Tabela 6, é visualizado os conjuntos de dados balanceados de acordo com o método do *undersampling*.

Tabela 6 – Distribuição das classes sociais nos conjuntos de dados balanceados

Cidade	Tabela do Censo	Forma de Rotulação	A	B	C	D	E	Total
Manaus	Pessoa	Centroide	104	104	104	104	104	520
		Setor predominante	107	107	107	107	107	535
	Responsável	Centroide	117	117	117	117	117	585
		Setor predominante	133	133	133	133	133	665
São Paulo	Pessoa	Centroide	141	141	141	141	141	705
		Setor predominante	142	142	142	142	142	710
	Responsável	Centroide	194	194	194	194	194	970
		Setor predominante	142	142	142	142	142	710

Fonte: Própria

A partir desses conjuntos de dados, foi separado a base de teste sendo 20% do total e a base de validação sendo 20% dos 80% restantes. Portanto, a base de dados de treino foi 64% dos conjuntos de dados totais.

Todos os treinamentos realizados foram utilizando o *Google Colabory* como ambiente de desenvolvimento tendo a GPU padrão como NVIDIA T4 Tensor Core GPUs.

## 7 RESULTADOS E DISCUSSÕES

O treinamento dos 16 modelos sendo 8 modelos da cidade de Manaus (AM) e 8 da cidade de São Paulo (SP), a próxima etapa é analisar os resultados dos modelos. Para isso, utilizou-se de métricas de avaliação dos modelos para problemas de classificação apresentadas na Seção 5.4.4, são elas: a precisão, revocação, F1-score e matriz de confusão. A partir das métricas, foram selecionados os modelos com desempenho melhores para uma análise mais detalhada.

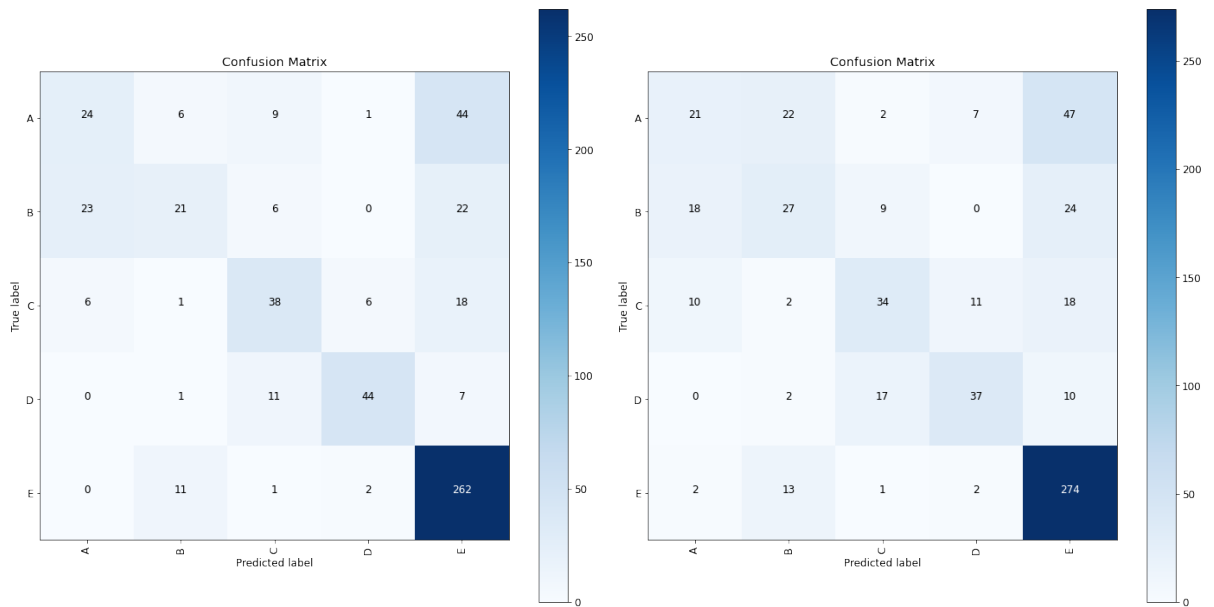
### 7.1 MATRIZ DE CONFUSÃO

A matriz de confusão é uma das métricas mais utilizadas para problemas de classificação como apresentada na seção 5.4.4.

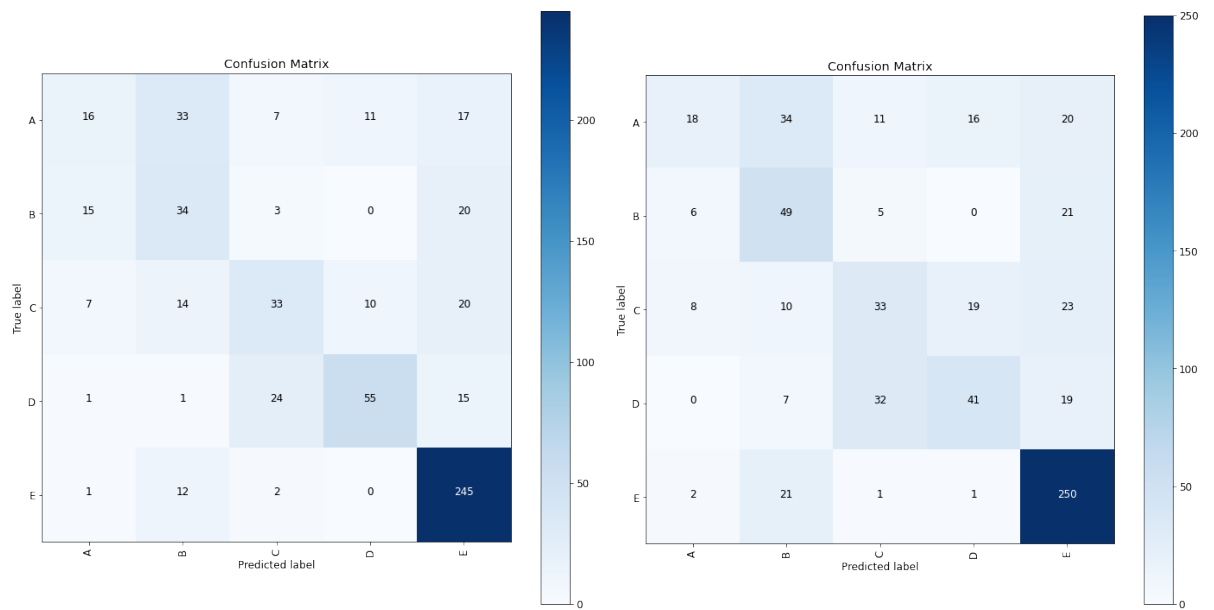
Na figura 21, são apresentadas as matrizes de confusão para os modelos de Manaus utilizando o treinamento com o aumento de dados e os pesos das categorias para a base de teste. O ideal da matriz de confusão é estar no sentido diagonal, ou seja, o modelo precisa ter a maior proporção de acertos em todas as classes presentes no conjunto de dados de teste, porém, é possível observar que todos os modelos tenderam para a classe E, visto que, é a classe social que mais possui imagens no conjunto de dados de Manaus, como visto na tabela 4. O modelo que mais possui a forma da diagonal é o que utilizou a tabela da Pessoa Renda do Censo de 2010 e o rótulo sendo o resultante da técnica do centroide no *grid* mencionada nas etapas anteriores.

Enquanto que, na figura 22, os modelos não estão mais tendendo para a classe E por conta da base de dados estar balanceada, ou seja, todas as classes têm o mesmo número de amostras. Neste caso, o modelo mais próximo do esperado foi considerando também a tabela Pessoa Renda, porém, pela técnica de rotulação das imagens pelo setor censitário predominante no *grid* da imagem. Isso porque é o modelo que mais está sabendo diferenciar as classes sociais extremas, a classe A e a classe E.

Figura 21 – Matrizes de confusão dos modelos da cidade de Manaus - Treinamento com o aumento de dados e pesos das categorias



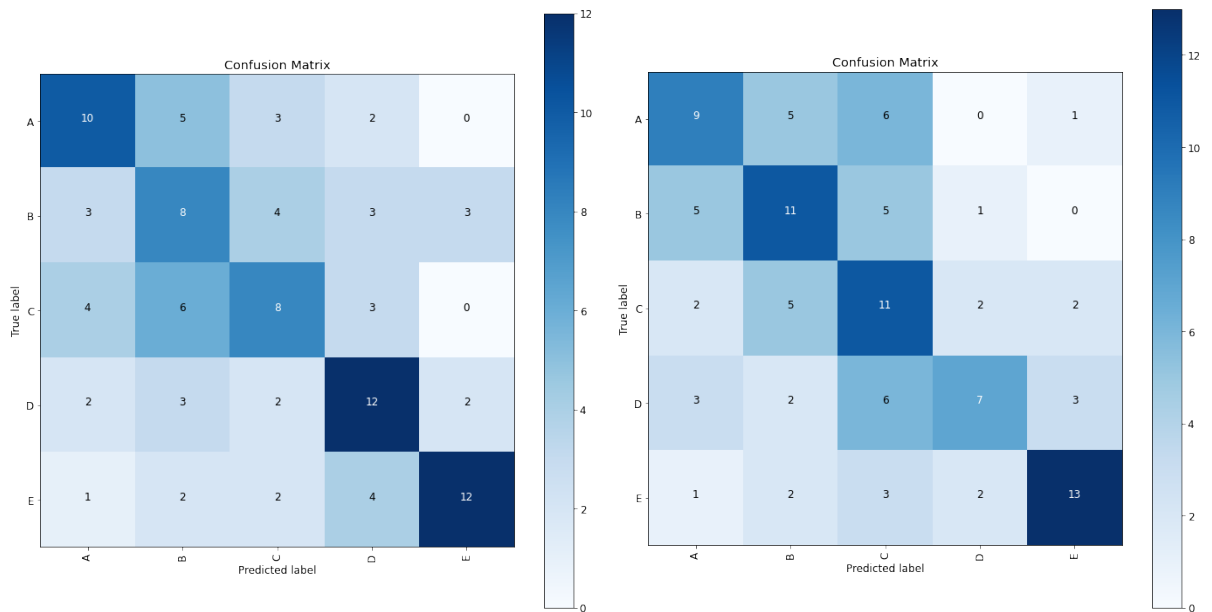
(a) Pessoa Renda - pelo centroide do *grid* da imagem (b) Pessoa Renda - pelo setor predominante no *grid* da imagem



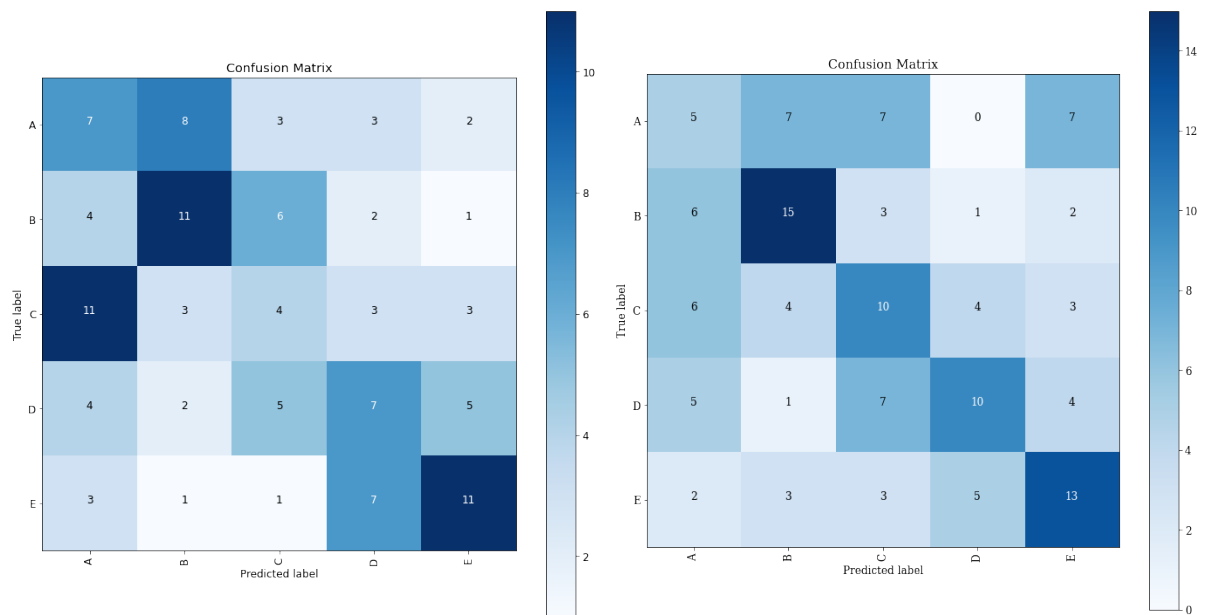
(c) Responsável Renda - pelo centroide do *grid* da imagem (d) Responsável Renda - pelo setor predominante no *grid* da imagem

Fonte: Própria

Figura 22 – Matrizes de confusão dos modelos da cidade de Manaus - Treinamento com o conjunto de dados balanceado



(a) Pessoa Renda - pelo centroide do *grid* da imagem (b) Pessoa Renda - pelo setor predominante no *grid* da imagem



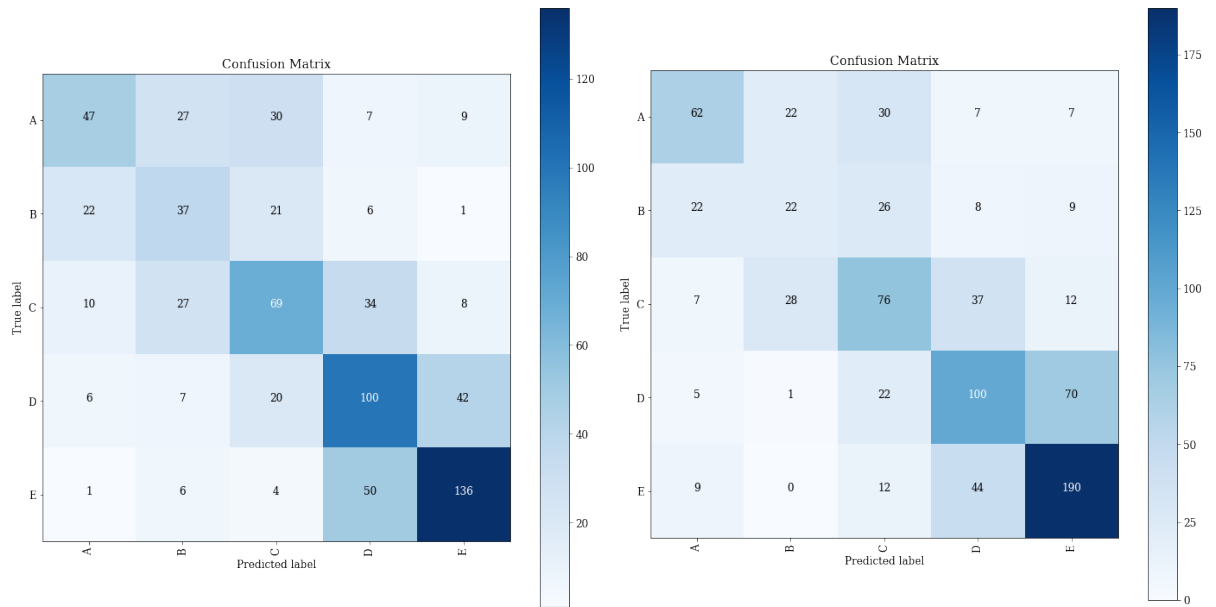
(c) Responsável Renda - pelo centroide do *grid* da imagem (d) Responsável Renda - pelo setor predominante no *grid* da imagem

Fonte: Própria

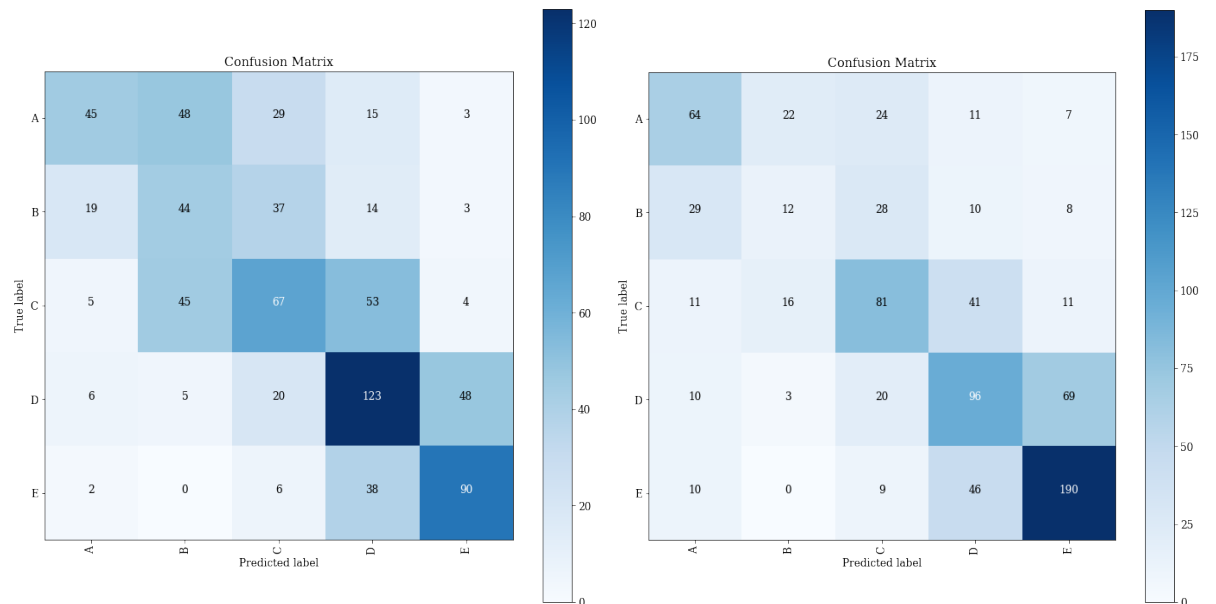
Nas figuras 23 e 24 são apresentadas as matrizes de confusão para a cidade de São Paulo. De forma geral, todos os modelos tiveram um desempenho melhor que os modelos de Manaus devido que, nos conjuntos de dados de São Paulo, as classes estão mais equilibradas do que nos conjuntos de dados de Manaus. Para o treinamento com o

aumento de dados e pesos das classes, o modelo mais adequado é o da figura 23a porque o mesmo consegue diferenciar as classes sociais altas (classes A e B) das classes sociais baixas (classes D e E), além de também conseguir acertar mais a classe B e a C.

Figura 23 – Matrizes de confusão dos modelos da cidade de São Paulo - Treinamento com o aumento de dados e pesos das categorias



(a) Pessoa Renda - pelo centroide do *grid* da imagem (b) Pessoa Renda - pelo setor predominante no *grid* da imagem

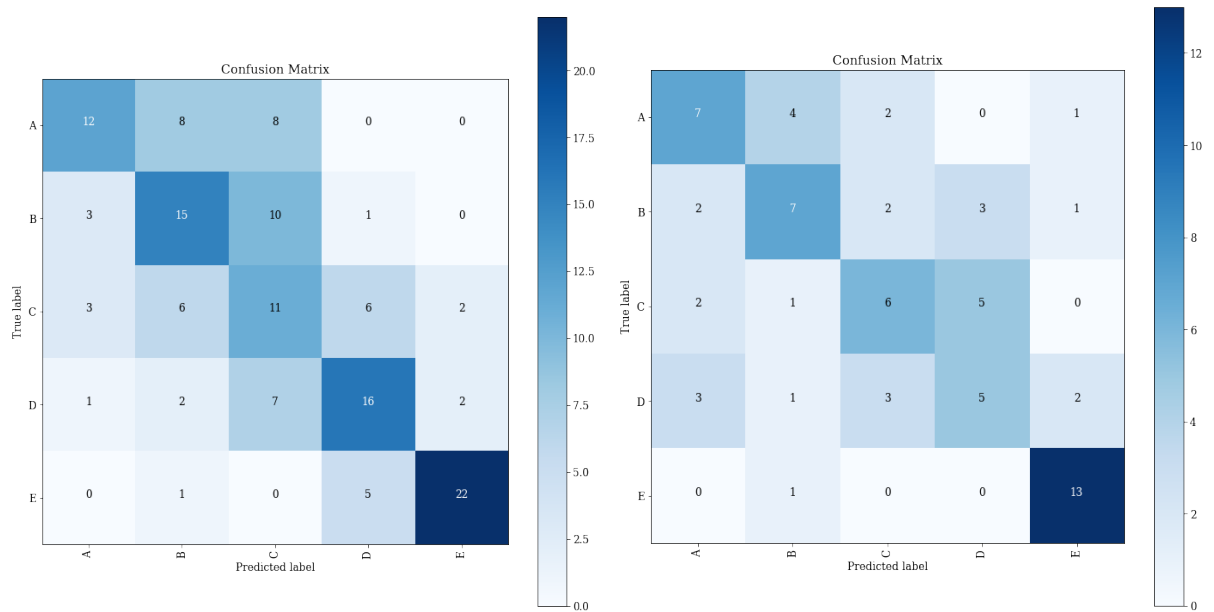


(c) Responsável Renda - pelo centroide do *grid* da imagem (d) Responsável Renda - pelo setor predominante no *grid* da imagem

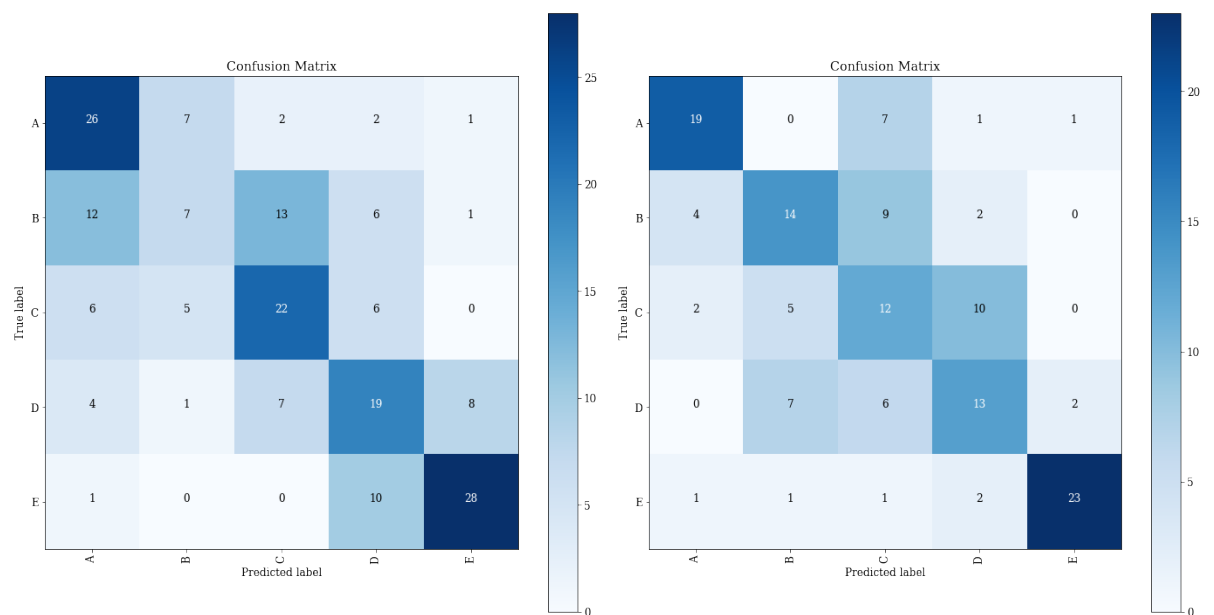
Fonte: Própria

Já para o treinamento com os conjuntos de dados balanceados, os modelos mais próximo do esperado são os das figuras 24c e 24d. O primeiro modelo é o que conseguiu diferenciar melhor a classe C, que é a classe média, enquanto que o segundo modelo conseguiu definir melhor a diagonal da matriz de confusão e, em ambos, foi possível diferenciar as classes sociais altas das classes sociais baixas.

Figura 24 – Matrizes de confusão dos modelos da cidade de São Paulo - Treinamento com o conjunto de dados balanceado



(a) Pessoa Renda - pelo centroide do *grid* da imagem (b) Pessoa Renda - pelo setor predominante no *grid* da imagem



(c) Responsável Renda - pelo centroide do *grid* da imagem (d) Responsável Renda - pelo setor predominante no *grid* da imagem

Fonte: Própria

## 7.2 PRECISÃO, REVOCAÇÃO E F1-SCORE

A partir da matriz de confusão, é possível calcular as métricas de avaliação de cada modelo nas bases de testes, que estão apresentadas nas tabelas 7, 8, 9 e 10. Os maiores valores das métricas estão em negrito para cada classe social e os modelos mais adequados levando em consideração as métricas e as matrizes de confusão estão em destaque.

Tabela 7 – Métricas de avaliação dos modelos de Manaus: Treinamento com o aumento de dados e pesos das categorias

Classes	Métricas	Pessoa Renda		Responsável Renda	
		Centroide*	Setor Predominante	Centroide	Setor Predominante
A	Precisão	0,45	0,41	0,40	<b>0,53</b>
	Revocação	<b>0,29</b>	0,21	0,19	0,18
	F1-Score	<b>0,35</b>	0,28	0,26	0,27
B	Precisão	<b>0,53</b>	0,41	0,36	0,40
	Revocação	0,29	0,35	0,47	<b>0,60</b>
	F1-Score	0,38	0,38	0,41	<b>0,49</b>
C	Precisão	<b>0,58</b>	0,54	0,48	0,40
	Revocação	<b>0,55</b>	0,45	0,39	0,35
	F1-Score	<b>0,57</b>	0,49	0,43	0,38
D	Precisão	<b>0,83</b>	0,65	0,72	0,53
	Revocação	<b>0,70</b>	0,56	0,57	0,41
	F1-Score	<b>0,76</b>	0,60	0,64	0,47
E	Precisão	0,74	0,73	<b>0,77</b>	0,75
	Revocação	<b>0,95</b>	0,94	0,94	0,91
	F1-Score	0,83	0,82	<b>0,85</b>	0,82
Geral	Precisão	<b>0,66</b>	0,61	0,62	0,59
	Revocação	<b>0,69</b>	0,64	0,64	0,60
	F1-Score	<b>0,66</b>	0,61	0,62	0,58

Fonte: Própria

Na tabela 7 são visualizadas as métricas de avaliação do modelo, cujo o treinamento foi utilizando o aumento de dados e os pesos das categorias, da cidade de Manaus. Analisando os valores das métricas e as matrizes de confusão da figura 21, nota-se que o melhor resultado geral foi de 0,66 de precisão, 0,69 de revocação e 0,66 de F1-Score utilizando a tabela Pessoa Renda do Censo de 2010 com a técnica de determinar a classe social da imagem sendo a classe do setor censitário que contem o centroide do *grid* da imagem.

Tabela 8 – Métricas de avaliação dos modelos de Manaus: Treinamento com as bases balanceadas

Classes	Métricas	Pessoa Renda		Responsável Renda	
		Centroide	Setor Predominante*	Centroide	Setor Predominante
A	Precisão	<b>0,50</b>	0,45	0,24	0,21
	Revocação	<b>0,50</b>	0,43	0,30	0,19
	F1-Score	<b>0,50</b>	0,44	0,27	0,20
B	Precisão	0,33	0,44	0,44	<b>0,50</b>
	Revocação	0,38	0,50	0,46	<b>0,56</b>
	F1-Score	0,36	0,47	0,45	<b>0,53</b>
C	Precisão	<b>0,42</b>	0,35	0,21	0,33
	Revocação	0,38	<b>0,50</b>	0,17	0,37
	F1-Score	0,40	<b>0,42</b>	0,19	0,35
D	Precisão	0,50	<b>0,58</b>	0,32	0,50
	Revocação	<b>0,57</b>	0,33	0,30	0,35
	F1-Score	<b>0,53</b>	0,42	0,31	0,43
E	Precisão	<b>0,71</b>	0,68	0,50	0,45
	Revocação	0,57	<b>0,62</b>	0,48	0,50
	F1-Score	0,63	<b>0,65</b>	0,49	0,47
Geral	Precisão	0,49	<b>0,50</b>	0,34	0,40
	Revocação	<b>0,48</b>	<b>0,48</b>	0,34	0,40
	F1-Score	<b>0,48</b>	<b>0,48</b>	0,34	0,40

Fonte: Própria

Já a tabela 8 apresenta as métricas de avaliação do modelo da cidade de Manaus, no qual o treinamento foi utilizando a base dados balanceada. Neste caso, uma maior parte das melhores métricas foram com a abordagem da tabela Pessoa Renda e com o método de verificar a classe social pelo centroide do *grid* da imagem com um F1-Score de 0,48, uma precisão de 0,49 e uma revocação de 0,48, porém, foi optado pelo modelo que foi treinado considerando o rótulo do setor predominante no *grid*, no qual possui a matriz de confusão da figura 22b, o valor do F1-Score sendo também de 0,48, a precisão de 0,50 e uma revocação também de 0,48. Isso devido ao modelo do centroide obter valores razoáveis para a maioria das classes sociais, mas para a classes social B obteve um resultado com valores de 0,33 para a precisão, 0,38 para a revocação e 0,36 para o F1-score, enquanto que o modelo utilizando a classe do setor predominante obtém um resultado mais balanceado entre todas as classes.



Tabela 9 – Métricas de avaliação dos modelos de São Paulo: Treinamento com o aumento de dados e pesos das categorias

Classes	Métricas	Pessoa Renda		Responsável Renda	
		Centroide*	Setor Predominante	Centroide	Setor Predominante
A	Precisão	0,55	<b>0,59</b>	0,58	0,52
	Revocação	0,39	0,48	0,32	<b>0,50</b>
	F1-Score	0,46	<b>0,53</b>	0,41	0,51
B	Precisão	<b>0,36</b>	0,30	0,31	0,23
	Revocação	<b>0,43</b>	0,25	0,38	0,14
	F1-Score	<b>0,39</b>	0,28	0,34	0,17
C	Precisão	0,48	0,46	0,42	<b>0,50</b>
	Revocação	0,47	0,47	0,39	<b>0,51</b>
	F1-Score	0,47	0,47	0,40	<b>0,50</b>
D	Precisão	<b>0,51</b>	<b>0,51</b>	<b>0,51</b>	0,47
	Revocação	0,57	0,51	<b>0,61</b>	0,48
	F1-Score	0,54	0,51	<b>0,55</b>	0,48
E	Precisão	<b>0,69</b>	0,66	0,61	0,67
	Revocação	0,69	<b>0,75</b>	0,66	<b>0,75</b>
	F1-Score	0,69	<b>0,70</b>	0,63	<b>0,70</b>
Geral	Precisão	<b>0,54</b>	<b>0,54</b>	0,49	0,52
	Revocação	<b>0,54</b>	<b>0,54</b>	0,48	<b>0,54</b>
	F1-Score	0,53	<b>0,54</b>	0,48	0,52

Fonte: Própria

A tabela 9 apresenta as métricas de avaliação do modelo da cidade de São Paulo, no qual o treinamento foi com o aumento de dados e com os pesos das categorias. As maiores métricas foram com a abordagem da tabela Pessoa Renda e do setor predominante no *grid* que possui 0,54 de F1-Score, precisão e revocação, porém, assim como no caso da tabela 8, a classe B também possui um resultado de 0,30 para a precisão, 0,25 para a revocação e 0,28 para o F1-Score, portanto, foi optado pelo método de considerar o rótulo da imagem pelo setor que contém o centroide do *grid* da imagem visto que possui a matriz de confusão mais próximo do desejado obtendo 0,54 para a precisão e revocação e 0,53 para o F1-Score no geral.

Tabela 10 – Métricas de avaliação dos modelos de São Paulo: Treinamento com as bases balanceadas

Classes	Métricas	Pessoa Renda		Responsável Renda	
		Centroide	Setor Predominante	Centroide	Setor Predominante*
A	Precisão	0,63	0,50	0,53	<b>0,73</b>
	Revocação	0,43	0,50	<b>0,68</b>	<b>0,68</b>
	F1-Score	0,51	0,50	0,60	<b>0,70</b>
B	Precisão	0,47	0,50	0,35	<b>0,52</b>
	Revocação	<b>0,52</b>	0,47	0,18	0,48
	F1-Score	0,49	0,48	0,24	<b>0,50</b>
C	Precisão	0,31	0,46	<b>0,50</b>	0,34
	Revocação	0,39	0,43	<b>0,56</b>	0,48
	F1-Score	0,34	0,44	<b>0,53</b>	0,38
D	Precisão	<b>0,57</b>	0,38	0,44	0,46
	Revocação	<b>0,57</b>	0,36	0,49	0,46
	F1-Score	<b>0,57</b>	0,37	0,46	0,46
E	Precisão	0,85	0,76	0,74	<b>0,88</b>
	Revocação	0,79	<b>0,93</b>	0,72	0,82
	F1-Score	0,81	0,84	0,73	<b>0,85</b>
Geral	Precisão	0,56	0,52	0,51	<b>0,59</b>
	Revocação	0,54	0,54	0,53	<b>0,57</b>
	F1-Score	0,55	0,53	0,51	<b>0,58</b>

Fonte: Própria

E, por último, a tabela 10 que apresenta as métricas de avaliação do modelo da cidade de São Paulo com o treinamento sendo com os dados balanceados. Neste caso, foi selecionado o modelo que utiliza a tabela do Responsável Renda com a abordagem de rotulação pelo setor predominante no *grid* devido que o mesmo possui as maiores métricas sendo 0,59 para a precisão, 0,57 para a revocação e 0,58 para o F1-Score considerando todas as classes sociais.

Contudo, os 2 modelos selecionados para a cidade de Manaus foram:

- a) Modelo 1: o treinamento com o aumento de dados de imagens e com os pesos das categorias presente no conjunto de dados, foi utilizando a tabela Pessoa Renda do Censo de 2010 com o método de rotulação sendo a classe social do setor censitário que contém o centroide do *grid* das imagens de satélite;
- b) Modelo 2: o treinamento com a base de dados estando balanceada, o melhor modelo foi utilizando a tabela Pessoa Renda com o método de rotulação sendo pelo setor

censitário predominante no *grid* da imagem de satélite.

Para os modelos de São Paulo, foram selecionados 2 modelos que possuíram melhores desempenho:

- a) Modelo 1: o treinamento com o aumento de dados de imagens e com os pesos das categorias presente no conjunto de dados, foi utilizando a tabela Pessoa Renda com o método de rotulação sendo a classe social do setor censitário que contém o centroide do *grid* das imagens de satélite;
- b) Modelo 2: treinado com a base de dados balanceada e sendo o rótulo através da tabela Responsável Renda e da técnica de rotulação através do setor predominante no *grid* da imagem.

Obtendo esses 4 modelos resultantes (2 de Manaus e 2 de São Paulo), foram feitas análises de como os modelos se comportam para as imagens de cada classe social, e assim, poder analisar os modelos de forma mais detalhada.

### 7.3 ANÁLISE DOS MODELOS EM CADA CLASSE SOCIAL

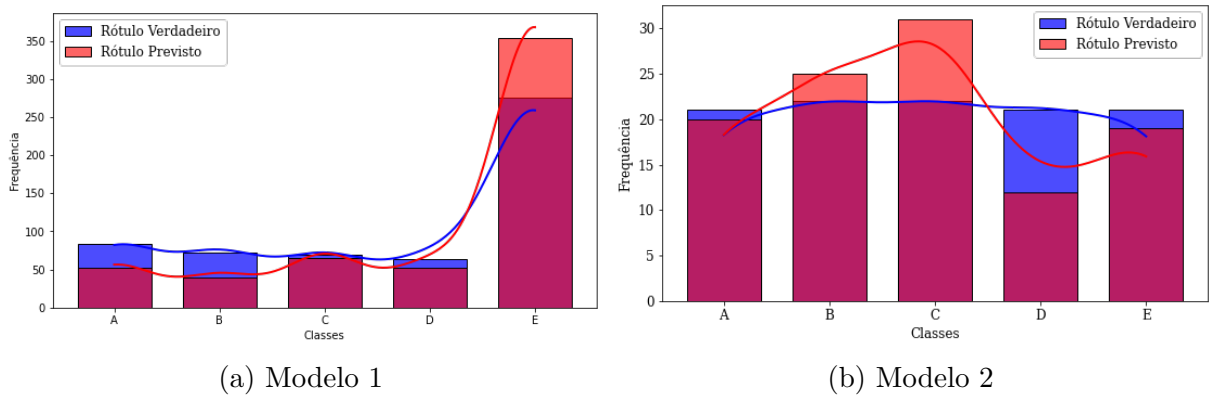
Nesta seção, é analisado o comportamento de cada modelo selecionado na etapa anterior em cada classe social. Todos os gráficos gerados e apresentados são os resultados dos modelos com os conjuntos de testes para cada abordagem.

#### 7.3.1 MODELOS SELECIONADOS DE MANAUS (AM)

Na figura 25 é apresentada a comparação dos rótulos verdadeiros das imagens com os rótulos previstos dos modelos na base de teste para comparar e verificar se os mesmos estão conseguindo acertar as classes sociais das imagens de satélite da cidade de Manaus.

No gráficos da figura 25a nota-se que a quantidade de imagens que o modelo afirmou ser da classe C é bem próximo do número de imagens que de fato é da classe C enquanto que, para a classe E, o modelo afirmou ter mais imagens da classe E do que realmente tem na base de teste, o que mostra que o mesmo tendeu para esta classe. Já para o gráfico do segundo modelo de Manaus que é representado na figura 25b, que é para o caso da base estar balanceada, ou seja, o número de imagens para cada classe é igual, é possível analisar que o modelo nesta abordagem tendeu um pouco para a classe C, porém, as outras classes estão bem próximas.

Figura 25 – Histogramas das classes sociais reais x das classes sociais previstas pelo modelo na base de teste - Manaus



Fonte: Própria

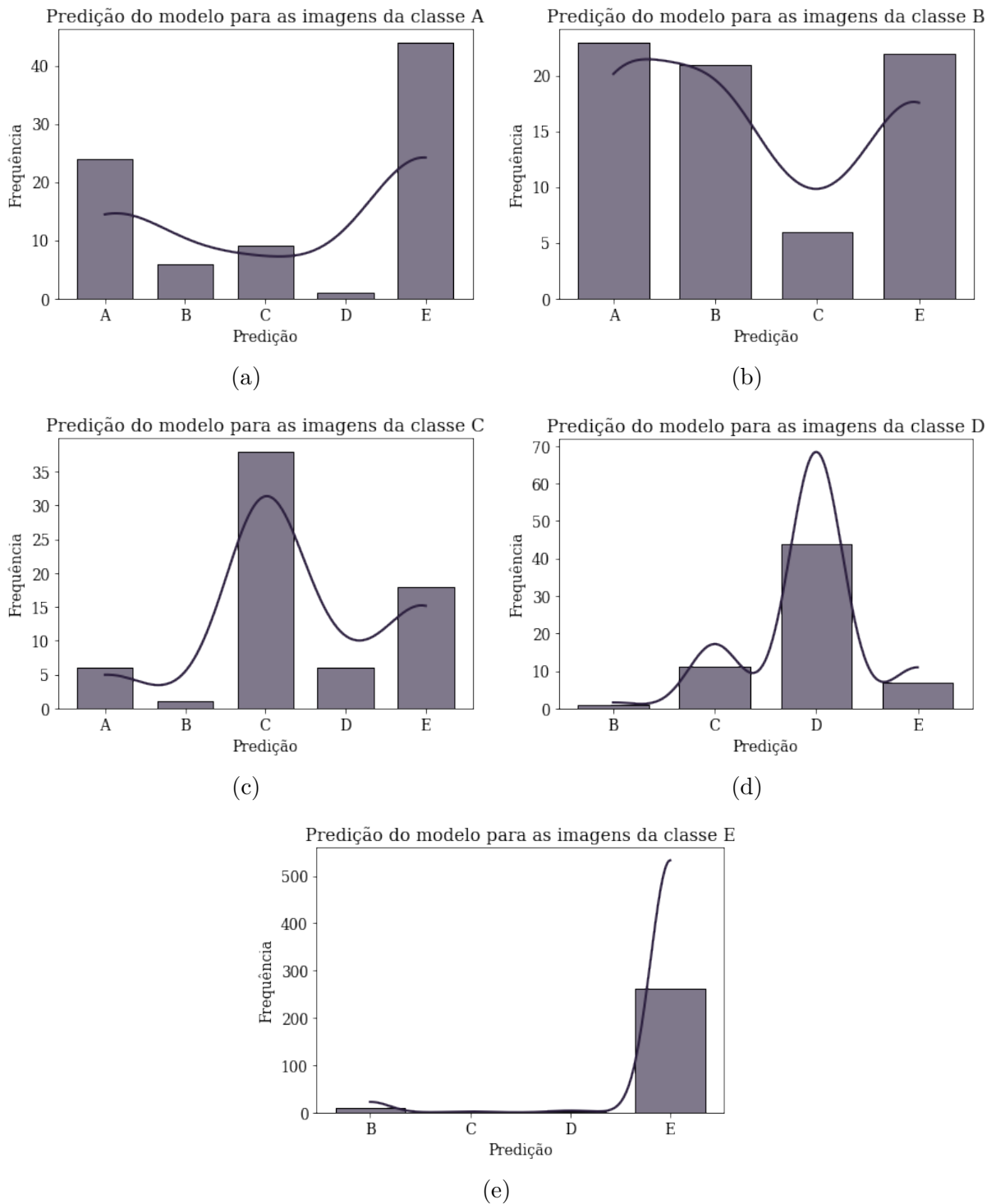
Como os gráficos da figura 25 apresenta somente a frequência real de cada classe social na base de teste e a frequência de cada classe social prevista pelo modelo, logo, não é possível definir os acertos dos modelos, visto que, os mesmos podem afirmar que uma imagem é da classe A, porém, é da classe B e, com isso, foram gerados os gráficos das figuras 26 e 27 e as tabelas 11 e Y com os valores dos erros absolutos para cada classe.

Na figura 26, é apresentado os histogramas com a distribuição de probabilidade para cada classe social do modelo 1, enquanto que a tabela 11 apresenta os Erros Médios Absolutos (MAE), onde que a equação é apresentado abaixo, para cada classe.

$$MAE = \frac{1}{n} * \sum_{i=1}^n |y_i - x_i| \quad (8)$$

Onde  $y_i$  é a predição em cada classe,  $x_i$  é o real de cada classe e o  $n$  é o tamanho da amostra. Para as imagens da classe A da base de teste, o modelo confundiu com a classe E sendo a classe de maior probabilidade seguida da própria classe A, isso pode ser observado pela tabela 11, onde que os maiores erros absolutos são da classe A, porque ele não acertou 60 imagens da base real da classe A, e da classe E com o erro sendo 44. Para as imagens da classe B, o modelo não soube diferenciar das classes A e E, onde que as probabilidades são próximas e pela tabela dos erros, essas classes são as que possuem os maiores erros absolutos. Já para as imagens da classe C, o modelo soube diferenciar das demais classes, o mesmo acontece para as imagens da classe D e da classe E. Contudo, este modelo obteve resultados não promissores para as classes A e B, no qual os Erros Médios Absolutos de 24 e 20,4, respectivamente, porém, obteve resultados esperados para as imagens das classes C, D e E com os erros médios sendo 12,4, 7,6 e 5,6, respectivamente.

Figura 26 – Histogramas com as distribuições de probabilidade para cada classe social do Modelo 1 da cidade de Manaus



Fonte: Própria

Tabela 11 – Erro Absoluto do Modelo 1 da cidade de Manaus

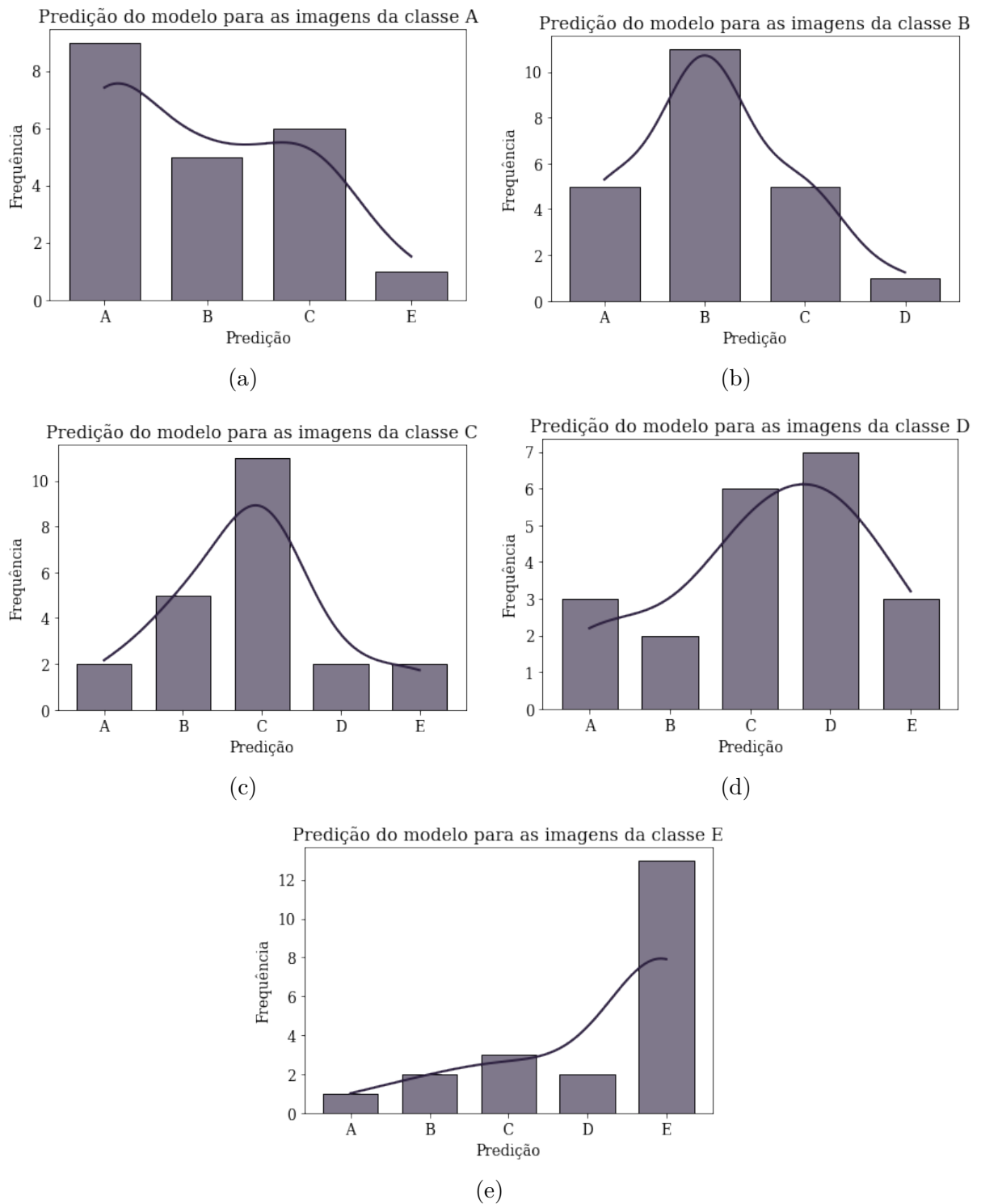
Classes Real	Erros Absolutos					Erro Médio Absoluto
	A	B	C	D	E	
A	60	6	9	1	44	24
B	23	51	6	0	22	20,4
C	6	1	31	6	18	12,4
D	0	1	11	19	7	7,6
E	0	11	1	2	14	5,6

Fonte: Própria

Já na figura 27, é apresentado os histogramas com a distribuição de probabilidade para cada classe social do modelo 2, enquanto que na tabela 12 apresenta os erros médios absolutos do modelo para cada classe.

Neste modelo, é analisado um comportamento interessante, onde que no gráficos das figuras 27b e 27c, que são das imagens das classes B e C, respectivamente, é observado que o modelo aprendeu as características dessas classes sabendo diferenciá-las, porém, o mesmo se confundiu com as classes vizinhas. Por exemplo, para as imagens da classe B, o modelo obteve a maior probabilidade para a própria classe e obteve a segunda e a terceira maior probabilidade para a classe A e C, que são as classes vizinhas da B. Isso pode ser analisado também pela tabela 12 que para as classes reais B e C é visto que os maiores erros absolutos são das classes vizinhas. Todavia, o modelo não obteve resultados adequados para as imagens das classes A e D podendo ser analisado pelos gráficos 27a e 27d, além disso, pela tabela 12 é notado que são os maiores erros médios absolutos sendo de 4,8 e 5,6, respectivamente.

Figura 27 – Histogramas com as distribuições de probabilidade para cada classe social do Modelo 2 da cidade de Manaus



Fonte: Própria

Tabela 12 – Erro Absoluto do Modelo 2 da cidade de Manaus

Classes Real	Erros Absolutos					Erro Médio Absoluto
	A	B	C	D	E	
A	12	5	6	0	1	4,8
B	5	11	5	1	0	4,4
C	2	5	11	2	2	4,4
D	3	2	6	14	3	5,6
E	1	2	3	2	8	3,2

Fonte: Própria

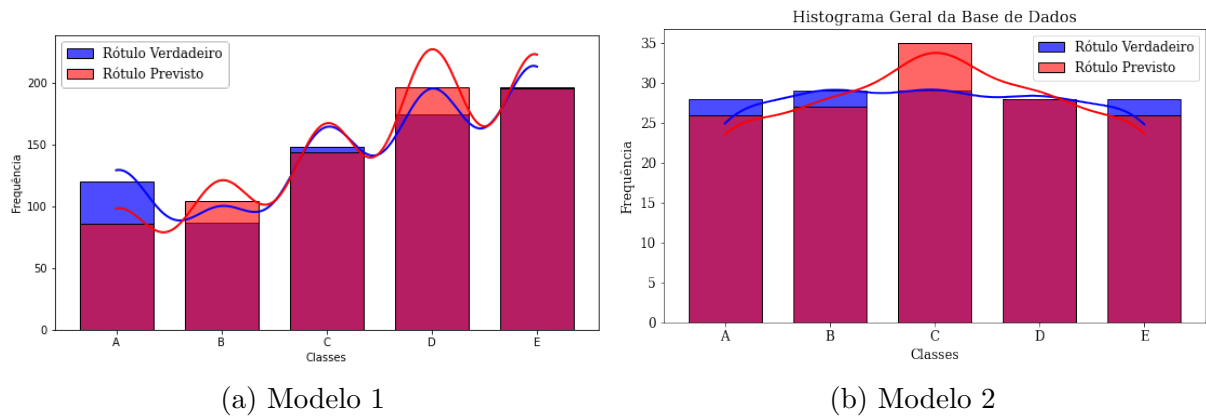
Analisando de forma geral, o modelo 2 obteve um desempenho mais próximo do esperado do que modelo 1, apesar do F1-Score para o modelo 1 ser de 0,66 e para o modelo 2 ser de 0,48. Isso pode ter sido por conta dos conjuntos de dados estarem muito desbalanceados para o caso do modelo 1, ou seja, tinha muitas imagens das classes D e E apresentando os menores erros médios absolutos e maiores métricas de precisão e revocação elevando o valor do F1-Score, porém não obtinha muitas imagens para as classes A e B, que já são as classes que tem os maiores erros médios absolutos e apresentam menores valores nas métricas de precisão e revocação, logo dificultou o modelo aprender características o suficiente para conseguir classificar as classes mais altas. Além disso, a cidade de Manaus é cercada por áreas verdes, o que pode ter dificultado para o modelo conseguir analisar as características de uma determinada área das imagens.

### 7.3.2 MODELOS SELECIONADOS DE SÃO PAULO (SP)

Os histogramas de comparação das classes reais da base de teste com as classes previstas pelos modelos são apresentados na figura 28. De forma geral, o número de imagens de cada classe prevista pelo modelo é próximo do número real para cada classe na base de teste, somente a classe B do gráfico da figura 28b que o número de imagens que o modelo afirmou ser da classe B é menor do que o número real, o que implica que a revocação deste modelo para a classe B é baixa como vimos na tabela 10.



Figura 28 – Histogramas das classes sociais reais x das classes sociais previstas pelo modelo na base de teste - São Paulo

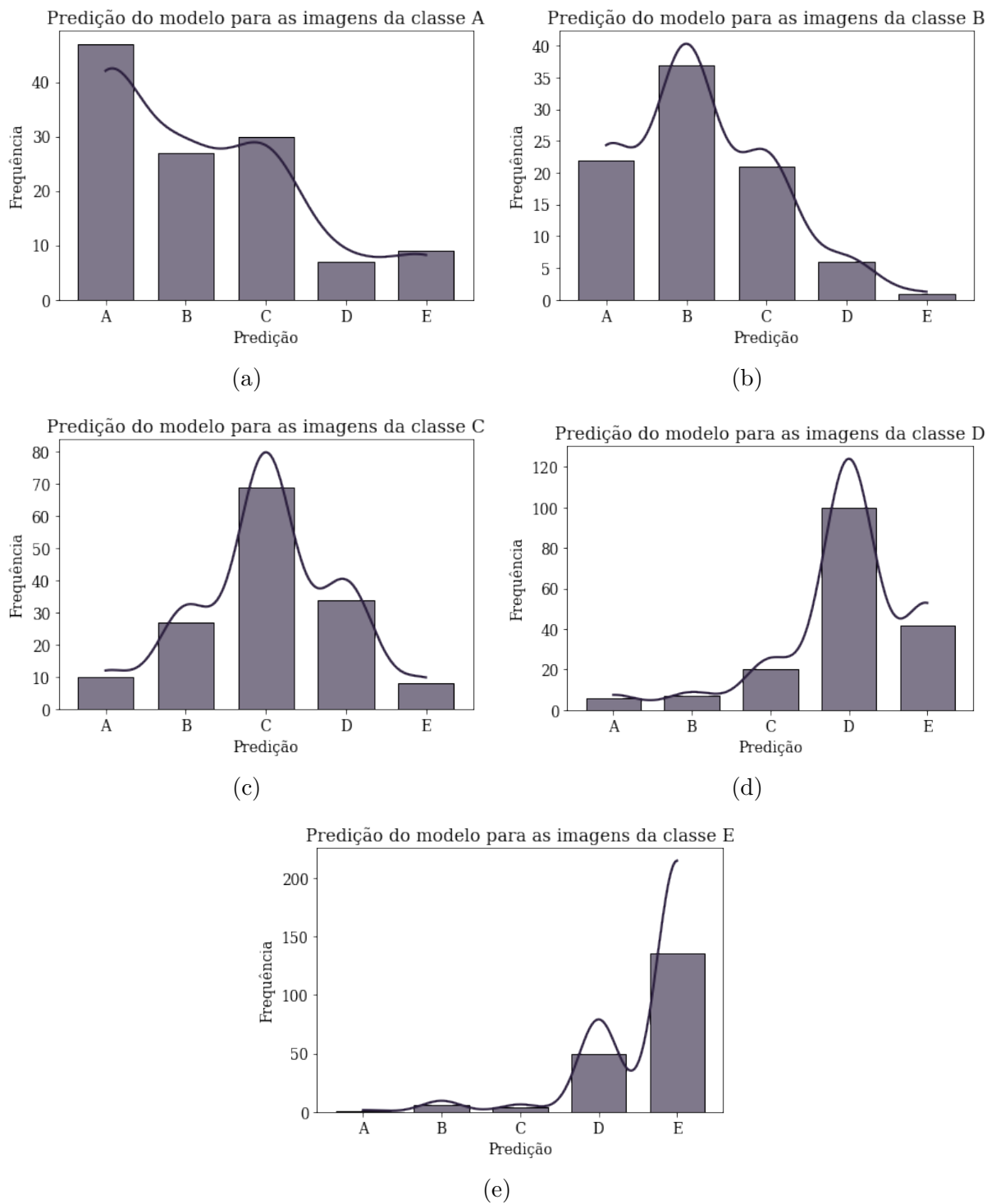


Fonte: Própria

Através dos histogramas do Modelo 1 apresentado na figura 29, percebe-se que para as imagens de todas as classes o modelo acertou para a maioria dos casos, ou seja, para as imagens da classe A da base de teste, o modelo afirmou grande parte dessas imagens sendo da classe A, para a classe B, C, D e E também a mesma situação, ou seja, a curva de distribuição de probabilidade fica com a média na respectiva classe.

Um ponto interessante é que para quase todas as classes, com exceção da classe A, a segunda e a terceira maior frequência foram de classes vizinhas, ou seja, o modelo confundiu a respectiva classe com as classes vizinhas, e isso também pode ser analisado na tabela 13, onde que os maiores erros absolutos em cada classe real são das classes vizinhas, o que pode-se concluir que em algumas situações as infraestruturas dos edifícios, ruas, etc., não possuem grandes diferenças entre as classes próximas e, assim, afetar o desempenho do modelo para acertar a devida classe. Por exemplo, as áreas da classe A podem ter alguma semelhança com as áreas da classe B nos casos quando as pessoas daquela região da imagem estão migrando da classe B para a A ou vice-versa, assim acontece também com a classe D e E e, para a classe C é entre B e D visto que fica no intermediário entre a classe alta e a classe baixa.

Figura 29 – Histogramas com as distribuições de probabilidade para cada classe social do Modelo 1 da cidade de São Paulo



Fonte: Própria

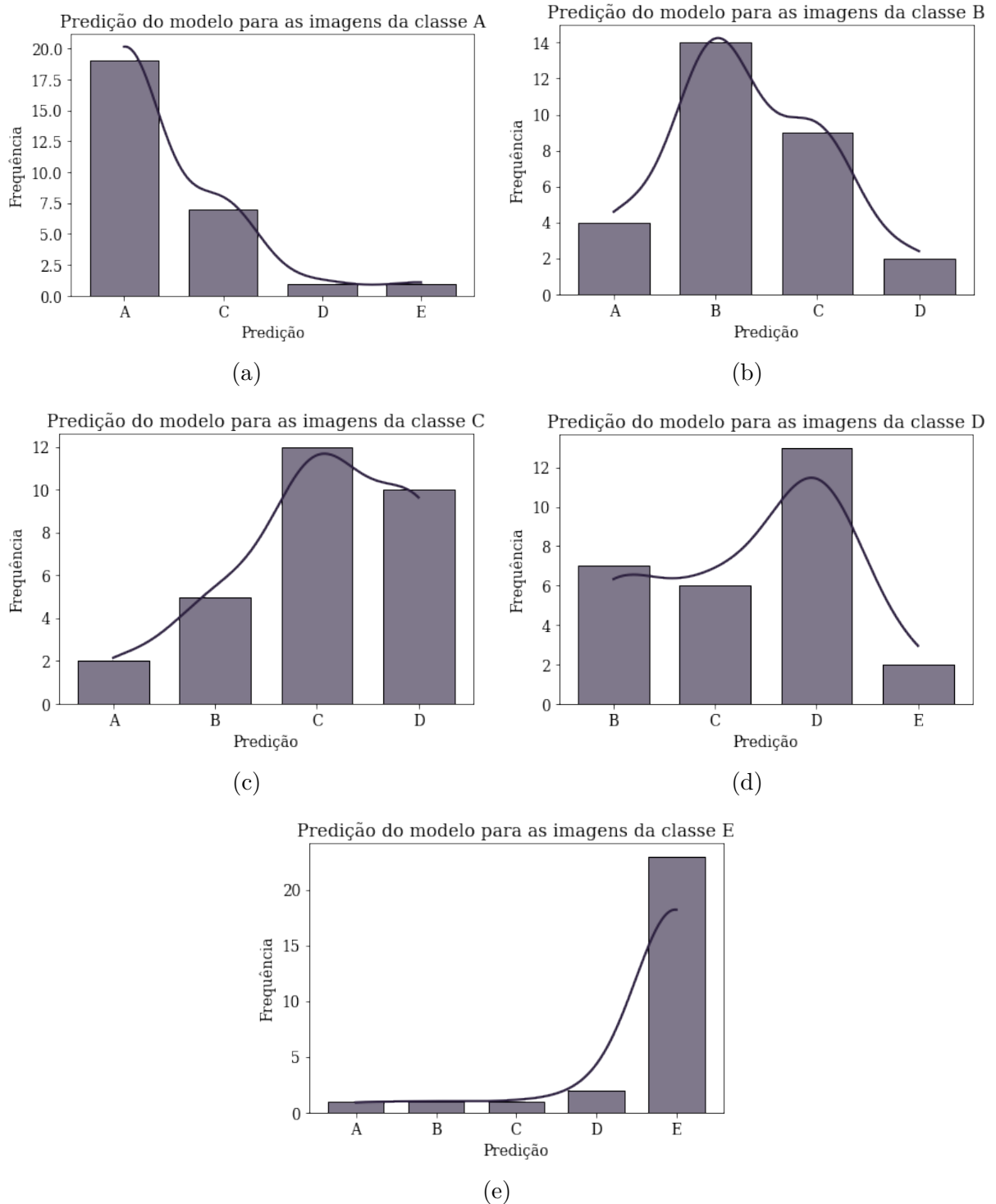
Tabela 13 – Erro Absoluto do Modelo 1 da cidade de São Paulo

Classes Real	Erros Absolutos					Erro Médio Absoluto
	A	B	C	D	E	
A	73	27	30	7	9	29,2
B	22	50	21	6	1	20
C	10	27	11	79	8	31,6
D	6	7	20	75	42	30
E	1	6	4	50	61	24,4

Fonte: Própria

Com os histogramas do modelo 2 de São Paulo apresentados na figura 30 e pela tabela 14 contendo os valores dos erros médios absolutos para cada classe real das imagens, percebe-se que o modelo conseguiu acertar mais imagens para cada classe correta, principalmente a classe A e E, além de que ambas possuem os menores erros médios absolutos. É notado o mesmo comportamento das classes vizinhas para as classes B, C e E nas figuras 30b, 30c e 30e, respectivamente, e na tabela 14 que os erros absolutos nas classes próximas às classes reais são maiores que nas classes distantes. Por exemplo, na classe real B, o erro absoluto com a classe E é igual à 0, o mesmo acontece para a classe real C. Logo, para classes distantes o modelo consegue diferenciar, mas para as classes próximas, o modelo pode se confundir.

Figura 30 – Histogramas com as distribuições de probabilidade para cada classe social do Modelo 2 da cidade de São Paulo



Fonte: Própria

Tabela 14 – Erro Absoluto do Modelo 2 da cidade de São Paulo

Classes Real	Erros Absolutos					Erro Médio Absoluto
	A	B	C	D	E	
A	9	0	7	1	1	3,6
B	4	15	9	2	0	6
C	2	5	17	10	0	6,8
D	0	7	6	15	2	6
E	1	1	1	2	5	2

Fonte: Própria

Analisando os gráficos gerados e as métricas de validação com os valores dos erros médios absolutos através dos dados de teste, ambos os modelos obtiveram os desempenhos próximos. Sendo o primeiro modelo com 0,54 de Precisão, 0,54 de revocação e 0,53 de F1-Score e o segundo com 0,59 de Precisão, 0,57 de revocação e 0,58 de F1-Score, porém, ambos apresentam os casos em que os modelos conseguem diferenciar as classes distantes e podem não acertar quando as classes são muito próximas. Com isso, pode-se analisar uma distribuição equivalente com as classes altas agrupando as classes A e B, classe intermediária sendo a classe C e as classes baixas agrupando as classes D e E e, assim, aumentar as métricas de Precisão, F1-Score e revocação e diminuir os valores dos erros absolutos entre as classes próximas e, como consequência, diminuir o valor do erro médio absoluto das classes.

Além disso, comparando os modelos de São Paulo com os de Manaus, os de São Paulo obteve melhores desempenhos. Isso foi devido que todos os conjuntos de dados construídos de acordo com os processos descritos nas etapas de criação das bases de dados serem mais equilibrados do que os de Manaus, que possuem mais imagens de satélite da classe E do que das demais classes em todas as abordagens de rotulação.

Os treinamentos dos modelos tanto de Manaus quanto de São Paulo tiveram uma duração de, aproximadamente, 2 horas. Enquanto que para realizar a predição na base de teste, o processamento das imagens foi de 1 minuto, aproximadamente, utilizando a mesma GPU do processo de treino. Como foi realizado o *transfer learning*, o período de treino não foi tão prolongado, visto que, treinando uma rede neural convolucional do zero poderia durar mais horas ou até mesmo dias.

## 8 CONSIDERAÇÕES FINAIS

As pesquisas responsáveis pela coleta de dados socioeconômicos são caras e demoradas, por exemplo, o Censo Demográfico brasileiro (principal pesquisa que coleta os dados estatístico de toda área nacional) tem o período de, no mínimo, 10 anos, o que dificulta a elaboração mais precisas nas formulações de políticas públicas. Contudo, este trabalho teve o intuito de implementar os modelos de aprendizado profundo com imagens de satélite e dados do Censo para classificar as áreas das cidades de Manaus (AM) e São Paulo (SP) de acordo com as classes sociais.

A metodologia aplicada consistiu em 4 etapas: criação do conjunto de dados; a preparação dos dados de treino, teste e validação; treinamento dos modelos; e, por último, analisar as métricas e os resultados dos modelos na base de teste.

Houve 3 processos para a criação da base de dados, sendo eles: coletar as imagens de satélite por meio do *Google Maps Static API*; analisar as tabelas de renda do Censo; e, por último, integrar as imagens de satélite coletadas com os dados de renda sendo os seus rótulos, visto que, para problemas de classificação, a entrada do modelo é a própria imagem e a saída esperada do modelo é o rótulo desta imagem. Com a finalização do processo de rotular as imagens de acordo com os dados de renda, resultou-se em 4 formas de rotular uma única imagem, então foi realizados os treinamentos e os testes para cada método de rotulação com a finalidade de analisar qual método é mais eficaz.

A preparação dos dados de treino, teste e validação para o modelo consistiu em analisar quais técnicas seriam utilizadas no momento do treinamento. Diante de pesquisas e estudos, foi optado por treinar os modelos com os conjuntos de dados desbalanceados, ou seja, números de amostras diferentes para cada classe social, porém, utilizando aumento de dados no *dataset* por conta de que algumas classes haviam poucas amostras. Além disso, também foi abordado a questão de realizar os treinamentos com a base dos dados balanceadas, isto é, mesmo número de amostra para cada classe social.

Na etapa do treinamento, foram realizados 8 modelos para cada cidade (no total 16 modelos) considerando as abordagens definidas tanto na criação do conjunto de dados quanto na preparação das bases de treino, teste e validação. O treinamento foi utilizando a técnica do *transfer learning* com a arquitetura *EfficientNetV2* e os pesos iniciais das camadas sendo do *ImageNet*. Para validar os modelos, utilizou-se as principais métricas de validação para o problema de classificação, como a matriz de confusão, precisão, revocação e o F1-Score. As métricas de todos os modelos foram anotadas nas tabelas 7, 8, 9 e 10 e as matrizes de confusão nas figuras 21, 22, 23 e 24.

Analisando de forma geral os modelos para as imagens de Manaus (AM), os modelos obtiveram resultados de F1-Score sendo 0,66 para o Modelo 1 e 0,48 para o Modelo 2 analisados na Seção 7.3.1. Apesar do valor do F1-Score ser maior no modelo 1, o mesmo

tendeu para a classe E em quase todas classes. Isso pode ter sido por conta dos conjuntos de dados estarem muito desbalanceados para os casos que não houveram o balanceamento, ou seja, tinha muitas imagens das classes D e E, porém não tinha muitas imagens para as classes A, B e C, o que dificultou o modelo aprender características o suficiente para conseguir classificar estas classes. Já o modelo 2, apresenta os erros médios absolutos mais baixos e o modelo acertou mais as classes não havendo a tendência à classe E.

Enquanto que, para a cidade de São Paulo, os modelos analisados na Seção 7.3.2 apresentam valores de 0,53 de F1-Score e 0,58 de F1-Score para os Modelos 1 e 2, respectivamente. Ambos apresentam casos em que os modelos conseguem diferenciar as classes distantes e podem não acertar quando as classes são muito próximas. Com isso, pode-se analisar uma distribuição equivalente com as classes altas sendo as classes A e B, classe média sendo a classe C e as classes baixas sendo as classes D e E e, assim, aumentar a métrica de F1-Score e diminuir os valores dos erros absolutos entre as classes próximas e, como consequência, diminuir o valor do erro médio absoluto das classes.

O atual trabalho utiliza dados de renda de 2010 para rotular as imagens de satélite de 2022, o que pode ter impactado nos resultados dos modelos desenvolvidos, principalmente para a cidade de Manaus que mudou muito nestes últimos 10 anos. Contudo, sugere-se como melhoria deste trabalho, utilizar os dados de renda do Censo de 2022, que até o presente momento não está disponibilizado, para rotular de forma mais correta as imagens de satélite, visto que, o *Google Maps Static API* retorna imagens do momento da requisição. Além disso, aplicar modelos e técnicas de *Deep Learning* mais avançados com a finalidade de extrair melhor as características das imagens e, a partir delas, poder classificar as imagens de forma mais eficaz.

## REFERÊNCIAS BIBLIOGRÁFICAS

- CASTRO, D. A.; ÁLVAREZ, M. A. Predicting socioeconomic indicators using transfer learning on imagery data: an application in brazil. *GeoJournal*, 2022.
- CHOLLET, F. *Deep Learning with Python*. [S.l.]: Manning, 2018.
- ELACHI, C.; ZYL, J. van. *Introduction to the Physics and Techniques of Remote Sensing*. Second edition. [S.l.]: A JOHN WILEY SONS, INC., 2006.
- FATEHKIA, M. et al. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science*, 2020.
- GIL, A. C. *Como elaborar projetos de pesquisa*. 6. ed. [S.l.]: Atlas, 2019.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.: s.n.], 2015.
- GUPTA, S.; TAN, M. *EfficientNet-EdgeTPU: Creating Accelerator-Optimized Neural Networks with AutoML*. 2019. Disponível em: <https://ai.googleblog.com/2019/08/efficientnet-edgetpu-creating.html>.
- HOWARD, A. G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017. Disponível em: <https://arxiv.org/abs/1704.04861>.
- IBGE, I. B. de Geografia e E. *Introdução ao Processamento Digital de Imagens*. [S.l.]: IBGE, 2001.
- IBGE, I. B. de Geografia e E. Base de informações do censo demográfico 2010: Resultados do universo por setor censitário. 2010.
- IBGE, I. B. de Geografia e E. *Censo Demográfico: O que é*. 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/22827-censo-2020-censo4.html?=&t=o-que-e>. Acesso em: 14/09/2022.
- IBGE, I. B. de Geografia e E. *Malha de Setores Censitários*. 2023. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html?=&t=o-que-e>. Acesso em: 27/02/2023.
- KERAS. *About Keras*. 2022. Disponível em: <https://keras.io/about/>. Acesso em: 14/09/2022.
- KHAN, S. et al. *A Guide to Convolutional Neural Networks for Computer Vision*. [S.l.]: Morgan & Claypool, 2018.
- LECUN, Y. et al. Gradient based learning applied to document recognition. *Proceedings of the IEEE*, November 1998.
- LUTZ, M. *Programming Python*. Fourth edition. [S.l.]: O'Reilly, 2010.
- MARSLAND, S. *Machine Learning: An Algorithmic Perspective*. Second edition. [S.l.]: CRC Press, 2015.
- OPENCV. *About OpenCV*. 2022. Disponível em: <https://opencv.org/about/>. Acesso em: 14/09/2022.
- PIAGGESI, S. et al. Predicting city poverty using satellite imagery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.



- SHEEHAN, E. et al. Predicting economic development using geolocated wikipedia articles. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining*, 2019.
- SRINATH, K. R. Python – the fastest growing programming language. *International Research Journal of Engineering and Technology (IRJET)*, p. 354–357, December 2017.
- TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. 2019. Disponível em: [⟨https://arxiv.org/abs/1905.11946⟩](https://arxiv.org/abs/1905.11946).
- TAN, M.; LE, Q. V. Efficientnetv2: Smaller models and faster training. 2021. Disponível em: [⟨https://arxiv.org/abs/2104.00298⟩](https://arxiv.org/abs/2104.00298).
- TENSORFLOW. *Introdução ao TensorFlow*. 2022. Disponível em: [⟨https://www.tensorflow.org/learn?hl=pt-br⟩](https://www.tensorflow.org/learn?hl=pt-br). Acesso em: 14/09/2022.
- WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big Data*, 2016. Disponível em: [⟨https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6⟩](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6).
- WU, P.; TAN, Y. Estimation of poverty based on remote sensing image and convolutional neural network. *Advances in Remote Sensing*, 2019.
- ZAFAR, I. et al. *Hands On: Convolutional Neural Network with Tensorflow*. [S.l.]: Packt, 2018.
- ZHUANG, F. et al. A comprehensive survey on transfer learning. 2019. Disponível em: [⟨https://arxiv.org/abs/1911.02685⟩](https://arxiv.org/abs/1911.02685).